A consensus of evidence: the role of SPI-M-O in the UK COVID-19 response

Graham F. Medley

Department of Global Health and Development
London School of Hygiene and Tropical Medicine
London WC1E 7HT
United Kingdom

graham.medley@lshtm.ac.uk

The COVID-19 epidemic posed a set of challenges for the science-policy interface that were, at least initially, novel for most participants. In the UK, the scientific advisory group for emergencies (SAGE) committee system has been developed to collate and present evidence to policy-makers (i.e. civil servants) and decision-makers (i.e. elected representatives) (SAGE, 2022). I chair the sub-committee of SAGE focussing on transmission dynamics (SPI-M-O), which is one of the streams of scientific evidence that is considered alongside other streams, such as virology, clinical medicine, behaviour and public health. Modelling of the transmission dynamics of infectious disease has special considerations in its relationship to policy development, and this commentary is a personal reflection on some of those issues and is not a group or official view. There is already much investigation and discussion about the performance and role of SAGE and modelling during the COVID-19 epidemic, including a public inquiry (UK Covid-19 Inquiry, 2022). I am also aware that SARS-CoV-2 continues to be transmitted at high levels, and that SAGE and SPI-M-O might be asked to provide future evidence and advice.

SPI-M-O is composed of independent (i.e. non-government) academics, who work with a secretariat (i.e. government employees) to generate evidence based on modelling for input to SAGE and thence advice for policy development. SAGE does not advise which policies are developed, but rather the scientific evidence and advice which might be taken into account as policy is developed. There is a strong distinction between those within and without government in terms of sharing of information and required code of conduct, so that passing information freely between the two arenas is impossible: there is a fence that has to be worked through. The main output from SPI-M-O is a consensus statement which addresses questions we were asked by policymakers, and issues that members raised as being potentially policy-relevant. The subject-specific content is from members, but it is largely written by civil servants for civil servants. The consensus statements are all available on the SAGE website. Note that research published by members of SPI-M-O does not constitute SPI-M-O output unless it is included in the consensus statements.

My role has been largely three-fold. First, to understand what was being asked of the models and modellers, and to consider whether we had the resources and data to provide an answer. Second, to support the committee in coming to a consensus view which is useful, but which does not hide the inherent uncertainty. Third, to ensure that the consensus is understood by non-modellers. The first and third roles are largely translation between the

different languages used on either side of the fence, whereas the second role was rife with jargon and subject specific shorthand.

In the following, I go through several issues that were critical in developing scientific evidence so that it could inform government policy choices. There has been some misunderstanding and misrepresentation of the SPI-M-O role which I hope to clarify. It is critically important to analyse and understand the response to COVID-19, but here I simply highlight the issues I think important rather than attempt a full investigation, which is better done by others.

Epidemics Are Dynamic
Perhaps the single biggest problem, especially in the first few months of the pandemic, was the apparent lack of appreciation that pandemics are dynamic, and interventions change their shape, but cannot stop them other than through global eradication which was not possible without extensive international coordination and cooperation. Epidemics evolve as a consequence of the interactions between biological and social processes. Biology determines such aspects as infectious periods, transmissibility, pathogenicity and the impact of previous infection (through the immune response). Social factors determine the rate of transmission through the amount of contact that people have, which in turn is influenced by individuals' knowledge and experience of the epidemic. As the epidemic progresses, it changes both the biological landscape (principally through acquired immunity), and the social context through the population's attitude and behaviour. Consequently, transmission patterns at any point in time are altered by the past, and the transmission happening now will greatly influence the future. These interactions and the dependence of the present on the past are why mathematical models are essential to understand and interpret the data.

Government policy influences epidemics through advice to the public, regulation, and technology provision (e.g. testing and vaccination), all of which alter people's behaviour. The policies enacted are a response to the epidemic, but also change the epidemic. This interaction, like the epidemic itself, is dynamic, so that the policy enacted at one time will change the epidemic and make the required or optimum policies in the future different. Being able to have some insight into how policy is going to work now, and how it is going to shape the future is the key role of models in the advisory process. Ideally, models are used to be able to look across the whole epidemic and develop a strategy, as much as to inform the choices at any one point in time. The May 2020 strategy document has a picture of two epidemics in Figure 1 which does not have a timescale, but gives the impression that the measures taken in March 2020 had stopped the epidemic which is misleading as future waves were inevitable (HMG, 2020).

Given the role of policy in determining the epidemic, models must assume future policy choices: a key role of models is scenario analyses of "what if" different polices are used. The policy option to "do nothing", i.e. allow the epidemic to progress without government intervention, is a strong assumption. It is much better, generally, to have some insight what policy wishes to achieve so that models can be directed to inform specific choices. But up until about December 2020, there was relatively little discussion about the dynamics of the epidemic.

Science vs Policy

Ideally, science is a process of developing hypotheses and disproving them. In practice, politics is a process of deciding a course of action, then proving that the decision was correct. Scientists are necessarily trying to change knowledge, but politicians rarely admit to being wrong. These two approaches to understanding and improving the world we live in are very, very different. Decisions taken by politicians are necessarily political, but (ideally) based on the scientific understanding at the time, i.e. evidence-based. As science changes, so policy can evolve. In non-emergency situations, this development can be presented publicly as progress, and largely occurs "behind the scenes". The pivotal role for policy-makers (i.e. civil servants) is to translate between evidence and policy, i.e. between the very different realms of science and politics.

But in emergencies the public scrutiny of evidence, policy and decision is more intense. Even then, most scientific evidence can be presented as originating from outside of the policy arena. For example, drugs and vaccine policies naturally change as clinical trials report and clinical knowledge accrues. Evidence from models is different in that its creation requires assumptions about policy. In order to be able to talk across the fence, the modellers and policy-makers need a working relationship, built on trust. Whilst it is understandable that policy-makers do not wish to openly discuss all the policy options, the quality of evidence from modelling is improved the closer the relationship. As the epidemic progressed, the relationship developed, so that by January 2021 the modelling evidence could be directed towards informing the policy choices that fell within government's feasible territory. Prior to that, without informed insight into the overall strategy, modelling had to make assumptions about what a policy might be or when decisions might be made and implemented. For example Davies et al had to essentially guess what policy might be in order to be able to look across the epidemic (Davies et al., 2020). Models need questions to answer, otherwise they are just lines on graphs. If models are to inform policy choices and policy decisions then interaction throughout the process is essential.

Evidence vs Advice

The policy decisions during the COVID-19 epidemic had huge impact and consequence on the public's lives and deaths. These decisions should be made considering the best available evidence, which is the point of SAGE. The modelling is part of that evidence and feeds into the SAGE consensus. Evidence is not advice – a distinction made clearly by Hine in the aftermath of the 2009 swine 'flu epidemic (Hine, 2010). Evidence can and should be public, but advice includes subjective interpretation of the evidence. In an epidemic, there is always a balance to be struck between reducing the rate of transmission and the cost of that reduction, where in the case of COVID-19, the cost was curtailing normal lives as much as economic. This balance is outside of the remit of scientific investigation because it involves value judgements: everybody has a different rate of exchange when trading jobs, freedom and health against each other.

A common question at the start of an epidemic is: what is the optimum policy? When the epidemic is over, what will have been a good, or even the best, outcome? In modelling terms, this defines an "objective function", i.e. what are we trying to minimise or maximise? Objective functions can be very complex, in which the trade-offs between different impacts are made explicit, or very simple, which is easier to communicate. In the event, politicians

are very reluctant to openly discuss the trade-offs (for good reasons perhaps), so that the models are asked to produce an array of outputs, which could then be combined with the outputs from separate economic forecasts for example. Whilst we were unable and not asked to do this during the epidemic, the combination of infectious disease dynamics with health economics is possible and highly informative for understanding what might be an optimal outcome (e.g. Tildesley et al. 2022).

Of course, there are multiple economic, legal, moral implications of policies and they cannot all be included in an objective function, so judgement is always required. Nonetheless, development of some broad guides on how to take models beyond the immediate health impact in future epidemics might be very useful. In terms of pandemic preparedness, a good outcome of the COVID-19 epidemic would be to have a clearer understanding of how policy-makers wish to cope with the multidimensional problems, including the inequalities that epidemics use and exacerbate. Personally, I believe that treating the dimensions separately is likely sub-optimal, and that better solutions might be found if, for example, economics and transmission dynamics are considered simultaneously. A key consideration for future epidemics is the time horizon before vaccines are available and the speed of delivery. The COVID-19 epidemic would have been very different if vaccines had not arrived until November 2021, or indeed had been available in summer 2020, and the overall strategy adopted needs to reflect the substantial change in context that vaccination brings.

The lack of a clear overall goal makes longer-term strategy development very difficult, and effectively confines models to providing evidence for the immediate, tactical decisions, as has happened in previous epidemics (Green & Medley, 2002). The time pressures for providing evidence during an epidemic effectively reduce the opportunity to be able to consider strategy, which emphasises the importance of modelling done between epidemics to learn the lessons of the last before the next starts. It is highly likely that all such consideration will support the conclusions that we cannot be over-prepared for epidemics, and that questions of strategy and general policy need to be answered before they start, and with the same urgency, commitment and resource as is given in the first few months after they start.

Evidence vs Prediction
The outputs of population models vary across the spectrum from quantitative forecasts to scenarios. The short-term forecasts make zero assumptions about future transmission, but consider the disease process as it unfolds in those already infected (SPI-M-Oa 2020, SPI-M-Ob 2020). These are what are usually termed predictions – they are quantifiable forecasts and can only be made for one or two generations of transmission into the future with any validity. At the start of the epidemic SPI-M-O produced such short-term forecasts for two weeks ahead, but by June 2020 most organisations using the forecasts had developed internal technology and techniques so their need from SAGE was superseded.

From September 2020 SPI-M-O started producing medium term projections (MTP), which required the strong assumption that transmission would continue as it was now for the next few weeks (SPI-M-Oc, 2020). We projected forward between 3 and 6 weeks, depending on how volatile current transmission was, and any impending changes to policy that we were

aware of. Whilst not formally predictions, these generally gave quantitative forecasts that proved reliable.

Generally, scenarios are not predictions. They make assumptions about the future to provide "what if" insights into potential patterns. We extended MTP into scenarios by making explicit assumptions about future transmission patterns, and unknown biological parameters (especially impact of vaccination). The MTP are scenarios in that they are "what if nothing changes for a few weeks", but because the first two weeks of disease outcomes are largely determined by the infections that have already occurred, they are a mixture of prediction with limited and specific scenarios.

Modelling teams on SPI-M-O also produced full transmission model scenarios, especially from January 2021 onwards, informing the potential patterns as vaccination was delivered and policy was changing at regular intervals on pre-announced dates. The value of scenarios is not lines on graphs, but comes from the synthesis of the insights that models produce which we produced as consensus statements, for example (SPI-M-O, 2021). Whilst the results include potential scales of outcomes (e.g. the relative size of impending waves), they more importantly highlight the important factors determining the future, such as vaccination coverage and the impact of variants. Whilst the common perception is that modelling produces graphs, it is the text of the SPI-M-O consensus statements that contains the most pertinent information to provide evidence for policy.

SPI-M-O also addressed questions about the effectiveness of interventions such as contact tracing, the formation of bubbles and use of rapid diagnostics. These were equally important for informing decisions but were less prominent, largely because they did not have the same relationship with policy, i.e. they do not have to assume future policy to be produced. The range and depth of what the members achieved is being published, for example Brooks-Pollock et al (2021), as well as in the extensive documentation on the SAGE website. Whilst it is always possible to argue that we had insufficient diversity of approaches to overcome framework-based uncertainty there were very few occasions in which I felt personally that we had not addressed the policy question.

The common aspect of all our work was that it informs policy choices. When a decision must be made, having the available evidence at the time is essential. Whilst this has elements of prediction, predicting is not the purpose. Accurate prediction of people's behaviour is impossible, so accurate prediction of future disease transmission beyond a couple of generations of transmission is impossible. Insights can be gleaned, but the purpose is not to say "this is what is going to happen", but say "this is the evidence to inform your thinking". Note also that the quality of a decision is not determined by the outcome. It is quite possible to make the best decision at the time which turns out to have been sub-optimal in hindsight, and equally possible that a decision which ignores available evidence turns out better than expectes. SPI-M-O concern is the evidence at the time of the decision.

Uncertainty
In any capacity in which technical evidence and advice are being generated for decision-makers, it is critical that the uncertainty inherent in the processes are passed to the decision-makers. The uncertainty should be part of the decision. There is always temptation

and pressure to reduce the displayed uncertainty to make the decisions easier, but this would turn advisors into advocates. Making the uncertainty clear means that decision-makers must address the proper elements of the decision – especially their degree of risk aversion, and to the extent to which they are prepared to be precautionary or mitigate against different outcomes.

Developing a consensus about uncertainty is not straightforward. My personal view is that it is more easily done by a group external to government who do not have any (or have less) preconceived or inferred ideas of what policy preferences are. Whilst it is possible to define what is meant by uncertainty, agreeing what is uncertain and couching the uncertainty in context so that it will be understood correctly by policy-makers and decision-makers is the job of professionals. The SPI-M-O secretariat were essential in this role. Capturing the spectrum of views and producing evidence that has enough precision to be useful and enough breadth to be accurate is the essence of what SPI-M-O did. For example, the evidence in early 2021 was that another wave of transmission was highly likely, but the impact of that wave on hospitalisation and death was very uncertain, largely due to the unknown (and unknowable) characteristics of the vaccine and future variants.

The need to be explicit about uncertainty also highlighted the importance of consensus. Predictions about what is going to happen are relatively cheap to make and require no special knowledge, but agreeing the weight to be put on evidence to inform a decision is very different. Expertise in a field is largely being able to define the inaccuracy of the knowledge, and developing a consensus across a spectrum of expertise that focusses on the uncertainty was a key role for SPI-M-O. Given the time pressures and the lack of experience in developing evidence over such a long period and the lack of experience and knowledge of SARS-CoV-2, consensus would always be more useful than individual views.

The uncertainty and different preferences also make formally evaluating decisions after they are made problematic. The outcome is one measure of a decision, but bad decisions can produce good outcomes, and *vice versa*, especially given both the uncertainty, and that the definition of "good outcome" is rarely explicit. Whilst modelling can be used to create past counterfactuals (i.e. "what if we had done this instead"), given the time pressures and focus on the next decision, SPI-M-O did not generate such counterfactuals. For example, there has been much discussion about the timing of the first "lockdown" on 23 March 2020, with many claiming it was too late. Science can have no role in determining whether it was too late. There is clear scientific evidence that introducing the same restrictions with the same effect earlier and imposed for the same period would have prevented more deaths during the first wave (i.e. up to some specified time horizon), but SPI-M-O has not produced a consensus on the relationship between timing and outcome. There is also clear scientific evidence that a smaller wave of infection would have made future waves more likely and larger. Modelling cannot formally evaluate the correctness of the timing (nor the decision itself), but can indicate the potential outcomes of different timing and decisions if useful.

Policy information vs Advocacy
The role of SPI-M-O feeding evidence into SAGE has been sometimes misunderstood or misrepresented especially during the political discussion surrounding a decision. In particular, the absolute necessity of having discussion with policy-makers as evidence is

developed is construed as policy-makers having undue influence over the evidence and evidence being generated to support a particular policy choice. The alternative criticism was that the evidence was being constructed by the modellers in order to achieve their favoured policy outcomes, i.e. modellers were advocating or arguing for a particular intervention. These criticisms are mutually exclusive, and neither is accurate. The members of SPI-M-O were all very aware that it was not our role to make or even influence the decisions that properly, and legally, rest with elected representatives. Scientists who become advocates for particular policy responses lose their impartiality when it comes to providing evidence.

These criticisms also highlight the importance of consensus both as a barrier against control of the modelling evidence by particular views, and as a protection to individual members. There is no clear scientifically valid optimum policy, because of the preferential weighting of different types of damage an epidemic causes, and consensus effectively limits individual preferences. Models are not objective in that the modellers decide which data to include and how the model is constructed. The literature is littered with models that presuppose their results in their assumptions. A critical role of consensus is to challenge those assumptions: "tell me why I'm wrong" became part of the SPI-M-O jargon, often shortened to TMWIW. The peer-review function of model comparison and consensus cannot be underestimated.

As infectious disease modellers, the members of SPI-M-O have worked on a huge range of diseases in different circumstances, but being in the middle of a pandemic with direct and indirect impact on ourselves and our families was different. There was a shared aim of ensuring the decisions had the best information that we could possibly provide given the constraints, because we would be directly impacted by those decisions. None of us wanted to be subject to policy that was presented and chosen because of an individual's bias and personal circumstances. Under that pressure, the requirement to develop a consensus was also mutually supportive, and we all benefited from having a shared experience as the epidemic unfolded. I find it impossible to imagine producing the range of evidence, on short timescales, of such importance continuously for two years without being in a mutually supportive group with a shared purpose. In that respect, at least, I hope that SPI-M-O is seen by the members as completely successful.

Not public facing – evidence for policy choices
The role of SPI-M-O is to provide specialised evidence for SAGE, which then feeds into government decision-making. The role is not to inform the public, although given the importance of the decisions, the evidence that SPI-M-O produced quite rightly came under scrutiny. Modellers are very aware of the potential influence of their evidence – lines on graphs and multi-coloured maps can be very persuasive. Data are far more important than models – indeed models are a (complicated) transformation of the data. Models are useful to enhance understanding of the data, and certainly not to replace them, although it can be argued that no interpretation of the data is entirely model-free. SPI-M-O put a lot of effort into understanding the data available, especially its biases and delays. Again, the consensus approach – comparing and combining different data streams and different analyses of the same data – makes interpretation much more robust. Giving the modellers access to the same data was essential for consensus development.

The media and public interest created a complex set of challenges. There was a lot of emphasis on the members' independence, and being able to explain our work and role to the public and press is a demonstrable part of that independence. However, we are aware, that during an epidemic, clear communication is critical, and that epidemics have a bigger impact the less cohesive the response. Members also had responsibilities to their employers, who by and large were very supportive of giving their staff scope to contribute. Balancing these competing interests gave me, personally, some of the most stressful experiences of the epidemic.

Policy information vs Research
Policy timelines can be very short, especially given the timescale of the epidemic and the periods of rapid exponential growth. Evidence available after policies have been developed, or decisions have been made is of very limited value (Whitty, 2015). It is always possible to explicitly say nothing, but this is not satisfactory in the circumstances. Formal, scientific modelling was often not possible, so the challenge was to complete sufficient work to enable the available evidence to be assembled. Especially at early stages of the epidemic, SPI-M-O had many discussions which ended up with "we don't know, but this is our best guess" outcome. Again, this has to be balanced with the need to ensure that uncertainty is passed to decision-makers and that our guess is not misinterpreted as good evidence. Policy-makers also need to be told what better analyses and insight can be obtained with more time or more data. For example, the pace of the roadmap steps January-July 2021 was largely informed by the rate at which data could be accrued to understand the impact of each step.

Science works through competition, and the members of SPI-M-O are usually highly competitive. The UK is unusual in having such a large community of infectious disease modellers, which meant that there is a large array of expertise to call on in an emergency. Between epidemics there is intense competition for research funding, and even within the epidemic, groups set themselves the goal of publishing results before another groups. We had to avoid developing a collective view or "group-think", but in reality the alternative problem of ensuring that competition did not disrupt the development of a consensus was just as big an issue.

Conclusions
Consensus is important when producing evidence to support decisions in any situation, but during a pandemic becomes critical. The process of creating a consensus provides multiple functions: peer-review of methods and approach, emphasis of uncertainty, shared responsibility as protection to individuals and development of a shared purpose and the consequent emotional support. Whilst the process sometimes felt too slow and cumbersome, it is worth the outcome.

The interaction between modelling and policy, made more complicated by the politics, is a challenging context to work in. The translation of ideas and concepts between different areas of expertise needs to be rehearsed, and as Hine already found (see Recommendation 7), should not be left until it is needed. Having a well-drilled, well-resourced secretariat – the gatekeepers at the science-policy interface – is essential. The 2021 Civil Service Award to the SPI-M-O secretariat was very well deserved.

Evaluation of the role and performance of SPI-M-O as part of the SAGE and policy-making system is on-going. There are already several pieces of work that analyse aspects, for example Atkinson et al (2020), McCabe and Donnelly (2021) and Rhodes and Lancaster (2022). As in previous epidemics, the use of models to support longer-term strategy development, alongside the shorter-term tactical questions, was lacking, especially early 2020. My strong suspicion now is that it is very unlikely to happen in the immediate face of an epidemic, given the timescales and highly political setting. Developing strategies between epidemics is very important if policy is going to be able to respond rapidly at the start of an epidemic. It is much easier to present evidence about a nominated strategy than it is about all strategies. My hope is that pandemic preparedness remains high on the agenda until the next pandemic.

References

Atkinson, P., Gobat, N., Lant, S., Mableson, H., Pilbeam, C., Solomon, T., Tonkin-Crine, S., Sheard, S., 2020. Understanding the policy dynamics of COVID-19 in the UK: Early findings from interviews with policy makers and health care professionals. *Social Science & Medicine*, **266**, 113423, https://doi.org/10.1016/j.socscimed.2020.113423.

Brooks-Pollock, E., Danon, L., Jombart, T., Pellis, L., 2021. Modelling that shaped the early COVID-19 pandemic response in the UK. *Phil. Trans. R. Soc. B* **376,** 20210001. http://doi.org/10.1098/rstb.2021.0001

Davies N.G. et al, 2020. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health,* **5**, e375 - e385. https://doi.org/10.1016/S2468-2667(20)30133-X

Green, L.E., Medley, G.F., 2002. Mathematical modelling of the foot and mouth disease epidemic of 2001: strengths and weaknesses. *Research in Veterinary Science*, **73** (3), 201-205. https://doi.org/10.1016/S0034-5288(02)00106-6.

Hine, D., 2010. *The 2009 Influenza Pandemic – An independent review of the UK response to the 2009 influenza pandemic*. Cabinet Office, Crown Copyright. https://www.gov.uk/government/publications/independent-review-into-the-response-to-the-2009-swine-flu-pandemic (accessed 20 July 2022).

HMG, 2020. *Our Plan To Rebuild: The UK Government's COVID-19 recovery strategy*. ISBN 978-1-5286-1911-0, Crown Copyright. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/884760/Our_plan_to_rebuild_The_UK_Government_s_COVID-19_recovery_strategy.pdf (accessed 20 July 2022).

McCabe, R., Donnelly, C.A., 2021. Disease transmission and control modelling at the science–policy interface. *Interface Focus*, **11**, 20210013. https://doi.org/10.1098/rsfs.2021.0013.

Rhodes, T., Lancaster, K., 2022. Making pandemics big: On the situational performance of Covid-19 mathematical models. *Social Science & Medicine,* **301**, 114907, https://doi.org/10.1016/j.socscimed.2022.114907.

SAGE, 2022. https://www.gov.uk/government/organisations/scientific-advisory-group-for-emergencies (accessed 20 July 2022).

SPI-M-O, 2021. *Summary of modelling for scenarios for COVID-19 autumn and winter 2021-22.* https://www.gov.uk/government/publications/spi-m-o-summary-of-modelling-

for-scenarios-for-covid-19-autumn-and-winter-2021-to-2022-13-october-2021 (accessed 20 July 2022).

SPI-M-Oa, 2020. *COVID-19 short-term forecasting: Proposed process for discussion, 2 April 2020* https://www.gov.uk/government/publications/spi-m-o-covid-19-short-term-forecasting-proposed-process-for-discussion-2-april-2020 (accessed 20 July 2020).

SPI-M-Ob, 2020. *COVID-19 combined pilot model predictions, 30 March 2020.* https://www.gov.uk/government/publications/spi-m-o-covid-19-combined-pilot-model-predictions-30-march-2020 (accessed 20 July 2022).

SPI-M-Oc, 2020. *Medium-term projections and model descriptions.* https://www.gov.uk/government/publications/spi-m-o-covid-19-medium-term-projections-explainer-31-october-2020 (accessed 20 July 2022).

Tildesley, M.J., Vassall, A., Riley, S., Jit, M., Sandmann, F., Hill, E.M., Thompson, R.N., Atkins, B.D., Edmunds, W.J., Dyson, L., Keeling, M.J., 2022. Optimal health and economic impact of non-pharmaceutical intervention measures prior and post vaccination in England: a mathematical modelling study. *R. Soc. Open Sci*. 9: 211746, https://doi.org/10.1098/rsos.211746.

UK Covid-19 Inquiry. https://covid19.public-inquiry.uk (accessed 20 July 2022).

Whitty, C.J.M., 2015. What makes an academic paper useful for health policy?. *BMC Med* **13,** 301. https://doi.org/10.1186/s12916-015-0544-8