The Public Health Responsibility Deal: how should such a complex public health policy be evaluated?

Mark Petticrew, Elizabeth Eastmure, Nicholas Mays, Cecile Knai, Mary Alison Durand, Ellen Nolte

Policy Innovation Research Unit, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine (LSHTM), 15-17 Tavistock Place, London WC1H 9SH, UK
Address correspondence to Mark Petticrew, E-mail: mark.petticrew@lshtm.ac.uk

ABSTRACT

Background The Public Health Responsibility Deal (RD) in England was launched in 2011 as a public–private partnership which aims to ‘tap into the potential for businesses and other influential organisations to make a significant contribution to improving public health by helping us to create this environment’. It has come under criticism from public health advocates and others, who have suggested that it will be ineffective or perhaps even harmful. Like many public health policies, there have also been demands to know whether it ‘works’.

Methods We conducted a scoping review and used this, supplemented with interviews with stakeholders, to develop a detailed logic model of the RD (presented here) to help understand its likely outcomes and the pathways by which these may be achieved as a basis for planning an evaluation.

Conclusions Evaluations of complex interventions require not just assessment of effects (including outcomes), but also a clear conceptualization of the intervention and its processes. The way the RD and the pledges made by participant organizations has been presented makes it difficult at this stage to evaluate whether the RD ‘works’ in terms of improving health. Instead, any evaluation needs to put together a jigsaw of evidence about processes, mechanisms and potential future health and non-health impacts, in part using the current scientific evidence. This task is ongoing.

Keywords epidemiology, government and law, public health

Introduction: what is the Responsibility Deal?

The Public Health Responsibility Deal (RD) is a public–private partnership organized around a series of voluntary agreements that aims to bring together government, academic experts, and commercial and voluntary organizations to contribute to meeting public health objectives. Through it, businesses primarily, but also other organizations such as NHS Trusts and local authorities, commit to voluntary pledges to undertake actions for a public health benefit. The RD covers food, alcohol, physical activity, health at work and behaviour change. According to the former Secretary of State for Health, Andrew Lansley, under whom the RD was established, the aim is that ‘by working in partnership, public health, commercial and voluntary organisations can agree practical actions to secure more progress, more quickly, with less cost than legislation . . .’1 His successor, Jeremy Hunt, very recently emphasized his own commitment to the RD.2

The RD consists of core commitments (Box 1); supporting pledges, which define the operating principles and processes of the Deal (Box 2); and collective and organization-specific pledges. Collective pledges are collectively agreed actions (see Appendix 1). All partners (that is, businesses and other key organizations) are required to sign up to the core
Among their concerns was that the interests of industry had been prioritized over potential benefits to public health, and that no commitment had been made on alternative actions the Government would take if the pledges did not reduce alcohol-related harm (This predated the Government’s alcohol strategy published in March 2012). The House of Commons Health Select Committee was also not convinced that the ‘nudging’ approach exemplified by the RD would be effective. These concerns underline the importance of evaluating the RD robustly. However, one of the main challenges for its evaluation is that, like many complex interventions, it is not just one intervention. It is made up of many interacting components, operating at different levels, with different potential outcomes and mechanisms, implemented in very different contexts (food, alcohol, physical activity and health at work). The research questions in relation to any of these components, and the type and nature of evidence needed to determine whether and how any of these components work vary, and it is highly unlikely that there is a simple answer to the question, ‘Does it work?’. Another significant challenge lies in defining what ‘work’ really means. Referring to the stated RD objectives, which centre on tackling public health problems via a collaborative approach, many commentators will interpret these in terms of health outcomes. While these are obviously important, determining the effects on health may be impossible within any reasonable timescale, because of the lag between interventions and many population health effects. Moreover, there are other interpretations of ‘does it work’, which are also of importance. One of the key objectives of the RD was to bring a range of organizations with new and existing components, operating at different levels, with different potential outcomes and mechanisms, implemented in very different contexts (food, alcohol, physical activity and health at work).

Public health organizations have often been critical to the RD approach. Six public health organizations that were involved in the RD Alcohol Network publicly withdrew their support from the process before the RD was announced: Alcohol Concern, British Association for the Study of the Liver, British Liver Trust, British Medical Association, Institute of Alcohol Studies and the Royal College of Physicians. Among their concerns was that the interests of industry had been prioritized over potential benefits to public health, and that no commitment had been made on alternative actions the Government would take if the pledges did not reduce alcohol-related harm (This predated the Government’s alcohol strategy published in March 2012). The House of Commons Health Select Committee was also not convinced that the ‘nudging’ approach exemplified by the RD would be effective.

These concerns underline the importance of evaluating the RD robustly. However, one of the main challenges for its evaluation is that, like many complex interventions, it is not just one intervention. It is made up of many interacting components, operating at different levels, with different potential outcomes and mechanisms, implemented in very different contexts (food, alcohol, physical activity and health at work). The research questions in relation to any of these components, and the type and nature of evidence needed to determine whether and how any of these components work vary, and it is highly unlikely that there is a simple answer to the question, ‘Does it work?’. Another significant challenge lies in defining what ‘work’ really means. Referring to the stated RD objectives, which centre on tackling public health problems via a collaborative approach, many commentators will interpret these in terms of health outcomes. While these are obviously important, determining the effects on health may be impossible within any reasonable timescale, because of the lag between interventions and many population health effects. Moreover, there are other interpretations of ‘does it work’, which are also of importance. One of the key objectives of the RD was to bring a range of organizations with new and existing components, operating at different levels, with different potential outcomes and mechanisms, implemented in very different contexts (food, alcohol, physical activity and health at work).

Public health organizations have often been critical to the RD approach. Six public health organizations that were involved in the RD Alcohol Network publicly withdrew their support from the process before the RD was announced: Alcohol Concern, British Association for the Study of the Liver, British Liver Trust, British Medical Association, Institute of Alcohol Studies and the Royal College of Physicians. Among their concerns was that the interests of industry had been prioritized over potential benefits to public health, and that no commitment had been made on alternative actions the Government would take if the pledges did not reduce alcohol-related harm (This predated the Government’s alcohol strategy published in March 2012). The House of Commons Health Select Committee was also not convinced that the ‘nudging’ approach exemplified by the RD would be effective.

These concerns underline the importance of evaluating the RD robustly. However, one of the main challenges for its evaluation is that, like many complex interventions, it is not just one intervention. It is made up of many interacting components, operating at different levels, with different potential outcomes and mechanisms, implemented in very different contexts (food, alcohol, physical activity and health at work). The research questions in relation to any of these components, and the type and nature of evidence needed to determine whether and how any of these components work vary, and it is highly unlikely that there is a simple answer to the question, ‘Does it work?’. Another significant challenge lies in defining what ‘work’ really means. Referring to the stated RD objectives, which centre on tackling public health problems via a collaborative approach, many commentators will interpret these in terms of health outcomes. While these are obviously important, determining the effects on health may be impossible within any reasonable timescale, because of the lag between interventions and many population health effects. Moreover, there are other interpretations of ‘does it work’, which are also of importance. One of the key objectives of the RD was to bring a range of organizations with new and existing components, operating at different levels, with different potential outcomes and mechanisms, implemented in very different contexts (food, alcohol, physical activity and health at work).
business and government. The review included previous evaluations from any sector, and aimed to summarize the types of agreements that exist, how they worked in practice, the conditions for their success and how they had been evaluated. The intention was to understand what the processes and mechanisms underlying previous similar interventions had been, and what the outcomes had been. This would allow us to assess what research questions we needed to ask and what the challenges might be. The review would also inform the development of the logic model.

We also interviewed stakeholders, to develop a detailed logic model of the RD (presented here). This was intended as an aid to understanding its likely outcomes and the pathways by which these may be achieved, in order to act as a basis for planning an evaluation. Development of the logic model was also informed by interviews with five RD network chairs, and by analysis of policy documentation on the RD. The model itself was produced in Microsoft Visio (PC version).1

Results

The logic model: how is the RD expected to work?
The mechanisms by which the RD is expected to affect health are shown in the logic model (Fig. 1). This has not been done previously for the RD, and it was valuable in identifying the stages in its implementation and the data needed to assess whether progress is being achieved, in the absence of data on final health outcomes at this stage. In the case of the RD, it can be seen that the main pathway of activity runs from the initiation of the RD at the left of the diagram, through to the final outcomes on the right. The logic model also identifies activities/events and influencers that impact on the main activities. The formation of the Plenary Group is identified as the practical starting point for the intervention, which led to the formation of the five networks and to the development of the initial pledges. Subsequent stages along the pathway by which the RD is assumed to affect health include the negotiation and agreement of pledges, implementation of pledges by partners, the assumption that the implementation of the pledge results in a change of environment, which will lead to an improved health outcome for that individual, and finally, the assumption that the cumulative effect of the individual responses leads to a population-level health impact. The logic model also helps identify the key evaluation tasks at different stages (Table 1).

While the logic model will be revised and updated as more evaluation data are collected, note that most of the activities towards the centre of the logic model, rather than towards the right-hand end of the causal pathway, are more easily evaluated. Yet the right-hand side relates more closely to changes in behaviour and health status. As we move towards this side of the model, we have to rely increasingly on assumptions that pledges will have an eventual effect on health, but there may be less evidence of health outcomes. The focus of research is therefore initially likely to be on proof of concept and on evaluating processes, rather than on health outcomes.

Discussion

Main finding of this study

The development of the RD logic model showed that the health outcomes are clearly very important, but they lie at the end of a complex causal pathway which starts with engaging with business, and proceeds through the production of pledges whose outcomes could eventually have measurable impacts on health and health behaviours.

The key word here is ‘eventually’. The RD operates at two different levels—at the level of the Deal as a whole, and at the level of the individual pledges, and each level has different processes and outcomes, and needs different approaches to evaluation. The overall evaluation therefore needs to consider the operation of the RD, as well as the potential changes in knowledge, understanding and behaviour resulting from specific pledges. Evaluation of the RD as a whole needs to be oriented towards exploring whether the processes are in place to allow progress towards achieving health improvements, while evaluation of the pledges needs to be oriented towards determining whether they are achieving those health gains.

In terms of the RD overall, the many evaluation questions include the following:

What are the pros and cons of having the five networks together in one RD as opposed to being in separate initiatives? Are these the right networks? Are their members the right ones?

Is the RD exploited by individual businesses? For example, do some organizations avoid or delay actions that they would have been unable to avoid under legislation?

Is the RD likely to be faster, better and cheaper than the alternatives?

The research questions for each of the specific pledges are much more focused, such as assessing the likely impact on health-related outcomes, including consumption, of improving labelling of unit content of alcohol products. An economic perspective may also be of value. While it may not be possible to assess the cost-effectiveness of the RD as a whole, some formal assessment of the costs and benefits would be informative, and could include the costs of monitoring and the time involved in setting up and sustaining the Deal.
Partners commit to network pledges

Coalition Government is formed

Plenary Group is formed

First wave of pledges are proposed

Partners commit to core pledges and supporting pledges

Boards/Senior management of partners agree to commit to pledges

Partners are aware of the first wave of pledges

Partners are aware that competition law requirements will be met

Availability of alternative policy interventions

Public health benefit

Favourable publicity

Impact on profit

RD has credibility

RD has press coverage

Awareness of NGO involvement

Spin-off effects on reputation

Potential impact of competition or free-riding

Impact on market share

Potential for substitution or compensatory actions

Implementation plans are developed

Monitoring and evaluation plans are developed

Implementation of pledge results in a change in environment

Adherence over time

Individual has awareness of impact of pledge implementation

Individual change of behaviour

Improved individual level health outcome

Population level impact

Evidence of impact on behaviour change on health

Contribution and challenge of NGOs and academics

Development of ground rules for the researcher led approach

Encouraging and proposing innovation

Sufficient spread on

Evidence of behaviour change on health

Evidence of impact behaviour

Evidence of impact on individual

Evidence of impact on health

Supportive context

Development of ground rules for the researcher led approach

Encouraging and proposing innovation

Key to Logic model

Actions

Influencing factors

Key evaluation tasks

Fig. 1 The Responsibility Deal Logic Model.
Although they are obviously important, alternative approaches such as trying to assess the extent to which the RD as a whole has improved or is improving health are likely to be impossible, given the range of other influences on health which will also change over time.

**Evaluation of the individual pledges**

Evaluation of the individual pledges should yield information on whether each works in improving health. Ostensibly, this seems like a simple question to answer, and the causal pathways between many of the individual pledges and specific public health effects seem clear. However, the statement of a pledge by an organization does not in itself mean that that the effects of that pledge can be evaluated, or at least not yet. Each pledge can be seen as an intervention. It is widely understood that some interventions can be evaluated and some either cannot or should not, perhaps because it is not feasible to evaluate them, or because there is no clear research question that could be posed (for example, there may not be a clearly specified outcome relating to the pledge), or because it is simply too early in the developmental process of the intervention. Other pledges may be so limited in terms of their likely impact—or so distant from any direct health impact—that, given limited resources, they are not worth evaluation. Minor commitments to share information fall into this category.

A more systematic approach to determining whether it is feasible and meaningful to evaluate the individual pledges involves assessing whether they are sufficiently specific,
measurable and time-bound. Pledges that are not specific and measurable are so general that it is impossible to determine if and/or when their outcomes are achieved (e.g. a non-specific pledge would be one where the size of the anticipated change is not specified). Lack of a time dimension means that there is no point at which success or failure to achieve outcomes can be determined. Most of the pledges are not currently defined in ways that are amenable to evaluation: they are either not specific enough and/or not sufficiently health-related. This suggests that formative evaluation (to specify the pledges more closely) and process evaluation may be more important at this stage than evaluation of health outcomes.

Faster than regulation?
The previous Secretary of State for Health argued that the RD initiative would produce changes faster than the regulation could.1 Whether the results can indeed be achieved faster and better is not easy to answer because it depends on an imaginary counterfactual. It will be possible to compare implementing firms with non-implementing firms, but it should be remembered that the latter may not form a reliable counterfactual comparison. International comparisons are likely to be valuable in exploring the experience in other countries (e.g. in relation to introducing regulation or voluntary approaches), but again they may not be comparable in terms of government trying to achieve the same means via purely legislative means.

What is already known on this topic
From our scoping review it was clear that, if properly implemented and monitored, voluntary agreements can be an effective policy approach. However, it was equally clear that there is little evidence on whether they are more effective than compulsory approaches, and some of the most effective voluntary agreements have included substantial disincentives for non-participation and sanctions for non-compliance, which are absent from the RD. Many countries are moving towards these more formal approaches to voluntary agreements, which makes it important to understand not just whether they work, but also in what ways they ‘work’ or ‘do not work’.

What this study adds
What this pilot phase, and the resultant logic model, showed us is that, as a first step, for the RD to ‘work’, businesses have to deliver what they promise—so assessment of the details of the pledges and what they might mean for health is a key evaluation task, and involves assessing this systematically across all collective and individual pledges. Assessment of market penetration—that is, the extent to which the pledge affects a large enough proportion of consumers—is also crucial. Further necessary steps in the evaluation include, for specific pledges, assessing whether the action has an impact on consumers—for example, in the case of labelling, whether it is seen, understood and acted upon, and, if it is, whether it has a net effect on consumption of the item in question as part of the diet. Other effects, including effects on inequalities, adverse effects and compensation effects also need to be understood for individual pledges.

Limitations of this study
We were limited in that relatively few interviews were undertaken; however, these included the representatives of the main stakeholders, that is, the chairs of the RD networks. We also supplemented the interviews with discussions with relevant policy colleagues and by analysing the relevant policy documents. The main limitation of such a complex approach to the RD evaluation is that the decomposition of the ‘does it work?’ question into individual research tasks might obscure the simpler question of whether the RD is really of any value in public health terms. We do not believe this to be the case; rather, the unpacking of the causal pathways between the RD, and the individual pledges will allow a detailed and nuanced answer to the key questions of whether the RD works, and how.

Conclusions
It is simple to demand an evaluation of the RD. It is more complex to work out what can be evaluated, and how. It can be uncomfortable to recognize that not everything that seems to be important can be evaluated robustly. The evaluation of complex public health interventions, which operate at multiple levels and have multiple competing objectives, requires clear thinking about what can be evaluated and what types of evaluation can and should be done.

Timing is also crucial. The weakness of the public health evidence base is often criticized on the grounds that nothing appears to ‘work’. One possible reason for this is that interventions are frequently evaluated before they are fully formed or implemented; such evaluations thus almost inevitably produce negative or equivocal results. However, although this is a real risk, evaluators equally need to be wary of Buxton’s law from the field of health technology assessment: ‘It is always too early [for rigorous evaluation] until suddenly it’s too late.’11

Complex interventions require assessment, not just of effects (including outcomes), but also a clear conceptualization of the intervention and its processes. The evaluation of the RD therefore needs to put together a jigsaw of evidence
about processes, mechanisms and potential future health and non-health impacts, such as the knowledge, attitudes and behaviours of consumers, in part, using the scientific evidence we already have. This integrated approach will address multiple research questions, using a range of methods and data sources, and will ultimately shed light on the effectiveness or ineffectiveness of this voluntary, and controversial, approach to improving public health.

**Funding**

This work is funded by the Policy Research Programme of the Department of Health for England via its core support for the Policy Research Unit in Policy Innovation Research (PIRU, http://www.piru.ac.uk/). This is an independent report. The Department of Health had no role in writing the paper, and the views expressed are not necessarily those of the Department. We are not aware of conflicts of interest among members of the review team.

**References**