



## Review

## Do reviews of healthcare interventions teach us how to improve healthcare systems?

Ray Pawson<sup>a</sup>, Joanne Greenhalgh<sup>a,\*</sup>, Cathy Brennan<sup>b</sup>, Elizabeth Glidewell<sup>b</sup><sup>a</sup> School of Sociology and Social Policy, University of Leeds, Leeds, UK<sup>b</sup> Leeds Institute of Health Sciences, University of Leeds, Leeds, UK

## ARTICLE INFO

## Article history:

Received 3 October 2013

Received in revised form

6 March 2014

Accepted 20 May 2014

Available online 27 May 2014

## Keywords:

United Kingdom

Demand management

Health systems

Realist synthesis

Organisational change

Complexity

## ABSTRACT

Planners, managers and policy makers in modern health services are not without ingenuity – they will always try, try and try again. They face deep-seated or ‘wicked’ problems, which have complex roots in the labyrinthine structures through which healthcare is delivered. Accordingly, the interventions devised to deal with such stubborn problems usually come in the plural. Many different reforms are devised to deal with a particular stumbling block, which may be implemented sequentially, simultaneously or whenever policy fashion or funding dictates. This paper examines this predicament from the perspective of evidence based policy. How might researchers go about reviewing the evidence when they are faced with multiple or indeed competing interventions addressing the same problem? In the face of this plight a rather unheralded form of research synthesis has emerged, namely the ‘typological review’. We critically review the fortunes of this strategy. Separating the putative reforms into series of subtypes and producing a scorecard of their outcomes has the unintended effect of divorcing them all from an understanding of how organisations change. A more fruitful approach may lie in a ‘theory-driven review’ underpinned by an understanding of dynamics of social change in complex organisations. We test this thesis by examining the primary and secondary research on the many interventions designed to tackle a particularly wicked problem, namely the inexorable rise in demand for healthcare.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

## 1. Introduction

Modern healthcare systems are the epitome of complexity (Best et al., 2012; Clark, 2013; Thomson et al., 2012). This presents challenges to researchers who evaluate and synthesise interventions attempting to improve healthcare systems and to policy makers who implement their findings. In this paper, we question whether the labours involved in reviewing particular healthcare interventions inform us sufficiently about the business of improving healthcare systems.

To answer this question, we present findings from a realist synthesis of demand management (DM) interventions for planned care. All advanced health systems face substantial increases in activity and costs with a seemingly unstoppable rise in demand for all aspects of care. A particular strain is often felt on the matter of referral management (RM), where the patient is relayed from one part of the system to another, often without due care being given to

the balance of resources across the system. These are perennial problems and so potential solutions have been contemplated time and again and researched over and over again. In the case of DM/RM interventions we are confronted with a diversity of interventions, carrying an assortment of solutions, implemented to contrasting plans, distributed at quite different points in the health service structure. These interventions have been evaluated using a whole variety of different methods. In this paper, we draw on the complexity of demand management to illustrate some of the challenges in synthesising literature on health systems and argue that theory driven reviews offer a potential solution.

Discussion is organised in three sections. The first briefly describes the methodology of our review and we then present a sociologically informed model that recognises the whole-system character of organisation change, which we refer to as the ‘depth ontology’ of healthcare systems. The second section uses this model to examine organisational change as understood in the programme theories of the diverse interventions designed to harness demand and improve referrals. Here we discover a compartmentalised policy architecture in which separate interventions begin life by introducing reforms to specific parts of a healthcare

\* Corresponding author.

E-mail address: [j.greenhalgh@leeds.ac.uk](mailto:j.greenhalgh@leeds.ac.uk) (J. Greenhalgh).

service. We also observe that in the *longue durée* of implementation these interventions become more sensitive to the need for whole system change and thus revise our model accordingly. The third section examines the extent to which such system complexity has been addressed within existing reviews of referral management interventions. We find that evaluation research is almost entirely intervention based. Primary research thus tends to be marooned in the initial programme-theory compartments. This encourages a form of secondary research in which reviewers attempt to discover winners and losers amongst the rival types of sub-interventions. Whole system understanding is lost in the cycle of research production.

## 2. Realist synthesis

To frame the analysis presented in the rest of our paper, we begin with a brief exposition of our methodology, realist synthesis (Pawson, 2006). This methodology seeks to understand for whom and in what circumstances interventions work through identifying, testing and refining programme theories – the ideas and assumptions underlying how the various different interventions are intended to work. It aims to build explanation through the development of middle range theories (Merton, 1967), which remain close enough to the empirical phenomenon studied but which allow generalisations to be drawn. Programme theories guide the selection of middle range theories which can then be adapted and refined to serve as a framework for the rest of the review. We began by identifying the ‘programme theories’ underlying demand management interventions. These are most likely to be found in editorials, letters, think pieces and critical pieces and our search strategy reflected this (see Appendix 1). As we assembled these programme theories, it became clear that all DM/RM interventions seek to bring about organisational change. To generate a framework through which referral management programme theories and the associated empirical review literature could be understood, we developed a model of organisational change. In the remainder of this section, we describe the model and its origins in more detail. Our purpose here is not to provide a comprehensive review of complexity theory (Byrne and Callaghan, 2013) or organisational change theories (Greenhalgh et al., 2004; Murray et al., 2010), as these have been published elsewhere. Rather, our aim is to describe a model that provides an organising framework through which the change envisaged by referral management interventions can be explained and consequently, the existing reviews of referral management interventions can be critically interrogated in subsequent sections of the paper.

## 3. The depth ontology of organisational change

The nature of societal and organisational change has been explored for decades and our aim in this section is to extract one key precept, variously described as ‘depth ontology’, ‘vertical stratification’ and ‘systemic strain’ (Willmott, 2000; Zaltman et al., 1973). All of these concepts lay stress on the idea that organisations are layered and that change is mediated by intercommunication between these layers. We provide a highly abbreviated resume of this recurring theme using two sources which provide the basis for a simple model of the key, interrelated layers of change in health-care systems.

We begin with Hage and Aiken’s aptly titled book *Social Change in Complex Organisations* (Hage and Aiken, 1970). Sitting at the conceptual apex is something these authors refer to as ‘centralization’ of the organisation. This is the platform for core decision making; overarching strategies and systems are established here. Next comes ‘formalization’ – as the authors explain, ‘organizations

need daily guidelines for their operation; these guidelines are furnished by rules, the repository of past experiences’. At a level down, we have ‘stratification’, by which is meant the division of labour between members and the degree of differentiation in their roles. The next two strata are labelled ‘production’ and ‘efficiency’. These carry similar meaning referring to how organisational goals are embodied in everyday procedures and work routines. Finally in the model we reach the level of the individual member and the idea that change inheres in their ‘morale’, their levels of loyalty and team spirit.

All of these layers (systems, rules, roles, routines, morale) wrap and unwrap as institutions evolve. And it is this image which provides Hage and Aiken’s central thesis on the methodology for studying change: ‘A temporal perspective, one in which the focus of attention is not only on the interrelationships of parts of an organisation, but more importantly on changes in these interrelationships, is ultimately necessary for a more complete understanding of organizational behaviour’ (p. 28).

A rather grander attempt to understand the anatomy of societal change can be found in ‘realist’ methodology, most especially in the works of Archer (1995, 2003). Her theory of ‘morphogenesis’ attempts to answer the age old sociological chicken-or-egg question about what comes first in propelling social change – is it ‘structure’ or is it ‘agency’? Put simply, her answer is ‘chickenegg’. Peoples’ immediate actions are shaped within social structures in which they sit – communities, organisations, legal systems, power relationships, etc. However, in a longer time frame, these structures themselves change as a result of the activities and choices of the historically situated individuals who make them up. Social change, in short, occurs through a never ending cycle: ‘structural conditioning’ shapes ‘social interaction’ which in turn shapes ‘structural elaboration’, which then provides ‘structural conditioning’ and so on, and so on.

It takes more than structures and actors to make for morphogenesis and in the detail of Archer’s work other layers of social reality are identified – different ‘strata’ that are also real and are also causally powerful. Thus, for instance, she makes the distinction between ‘primary’ and ‘corporate’ agents: individuals with momentary, specific preferences and people with roles and positions who punch more heavily in their world shaping power. Similarly, structures have ‘first’ and ‘second’ order shaping powers such as that between a government and any administrative agency or quango it may create. Further, there are distinctions between structures in terms of how they exert influence, through the ‘material sphere’ of power and resources versus the ‘cultural sphere’ of information and know-how (Archer, 1995).

Many other authors have tried to uncover these vital constituents of social change (Hendry, 1996; Pettigrew et al., 1988). However, what is important is the overall recipe rather than the precise ingredients. Both models stress that it is the interdependency of the strata which enable and constrain change. They agree that each stratum has causal potential and that change can be instigated from any level. They posit that sustainable change requires harmonisation through the levels and that there is, of course, every reason to suspect that change provokes disharmony in institutions that are deeply stratified. Finally, these and similar models concur in supposing that change is something that no one steers. It happens perpetually and of its own accord whenever people and groups reflect on their own position.

This latter scenario raises an interesting conundrum – what happens when there is a formal intervention seeking system improvement? Can we expect navigation and control of the new ideas to be maintained across the system layers? This latter question is the supreme challenge for all investigations of service improvement programmes. To this end we provide an elementary model (Fig. 1) that helps to address it.

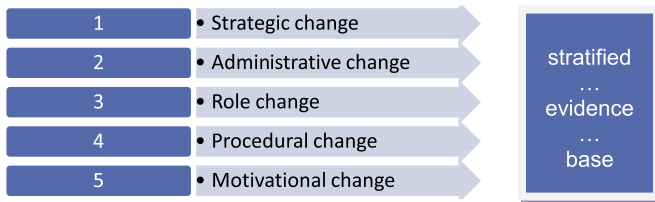


Fig. 1. Opportunities for change in health systems.

Figure one compresses health systems into just five strata, roughly corresponding to those discernible in the above literature. Each one is a potential arena of change, with interventions being directed at each of these layers. The first seeks change at the strategic level with plans to reconfigure and reorganise resources. The second modifies administrative and managerial responsibilities. The third rethinks the way in which individual roles and responsibilities are apportioned. The fourth type of change is task oriented, involving redesign in work routines and procedures. The fifth concentrates on individual members of organisations and seeks to improve their motivation and morale.

This structure exerts a remarkable influence on service redesign and healthcare inquiry. Opportunities for change tend to be devised at one or other level and programmes that follow gain their identity from the level targeted. Interventions commence within a given stratum and are then monitored and evaluated at that level. In the next section we test whether this model provides an adequate explanation of the fortunes of different referral management interventions within each strata through an analysis of the programme theories underlying these interventions.

#### 4. Organisational change as perceived in demand management interventions

As noted earlier, demand inflation and referral improvidence are perennial health service problems, which have gone on to inspire a medley of putative solutions. In this section we describe a spanning set of the major reforms. For each, we highlight typical examples of major UK NHS interventions, extracting brief illustrations of the core programme theory. Our task is to show that the initial modes of change envisaged sit squarely within the stratified model provided in Fig. 1. When planning change, programme architects grasp immediately for reforms of strategy or administration or roles or tasks or motivations.

Our analysis takes advantage of the longevity of DM/RM campaigns and also maps the evolution of these programme theories. We then consider the implications of these adaptations for our initial model of organisational change depicted in Fig. 1.

##### 4.1. Strategic change

The fundamental change envisaged in our first DM intervention involves rethinking the full length of the referral chain. The big idea is to improve the 'logistics' of the whole care pathway. This broad level of ambition fits with some of the grander, industrial-scale origins of the idea – 'lean thinking', 'queuing theory', 'continuous improvement', 'business process reengineering' (Erlang, 1909; Jones and Mitchell, 2006). In programme theory terms, the key change mechanism lies in the better coordination of practices and process across complex organisations. It is about getting the right provision in the right place, at the right time, in the right quantities. A useful illustration of the core idea is located in the following passage:

'...looking at the patient process and attempting to simplify and shorten it, identifying bottlenecks and pinch points for the individual process and then using the whole – hospital system perspective – to work out the best way of handing the flows into and through the process' (NHS Scotland, 2007b, p3).

Another significant feature of this family of intervention theories is their provenance. The grey literature comes in the form of 'reports' of major national agencies. Archetypical examples are the *Big Wizard* (NHS Modernisation Agency, 2002a) and the *Little Wizard* (NHS Modernisation Agency, 2002b). This strategic advice has a certain professional authority, deriving from what Hage and Aiken (1970) describe as 'centralization'. Although cast as advice on what to do in the hurly-burly of demand decisions, advice often follows glossy presentational formats associated with 'corporate strategy'. They are presented variously as 'toolkits', 'modules', 'frameworks', 'resources' and 'forums' – but above all as strategic 'recommendations'. For example, the *Little Wizard* (p9) proposes that "most delay, error and duplication is at the point where responsibility for the patient is handed from one person, department or agency to another" and therefore "Reducing the number of hand-offs can significantly reduce waits and delays". It also advises that "if possible, redesign the process to do all or some of the tasks simultaneously".

In summary the DM challenge lies in the entirety of the patient processing system. Less piecemeal and more strategic thinking is needed. What has history made of these ideas? Such recommendations often come externally and from on high from bodies like the Modernisation Agency. One obvious concern is about whether such missives might be ignored in the ward and the waiting room.

Accordingly, the underlying theory comes in for revision. The NHS Scotland report on *Patient Flow in Planned Care* (NHS Scotland, 2007a, p7) explains that: 'Improvements cannot take place without teams of people committed to improvement and with the knowledge required to change things' [*our italics*]. The report goes on to suggest the need for 'three generic team types that can be moulded and shaped depending on the improvement projects chosen' (NHS Scotland, 2007a). The trio is designated: 'Real Action Teams', 'Flow Improvement Teams', and 'Deep Analysis Teams' (p7). Here we see that programme theory is on the move. Strategic change (level 1) appears to rest upon organisational reengineering (level 2).

##### 4.2. Organisational change

The second proposed mechanism for improving DM/RM makes the case for transplanting one agency to do the work of another. The basic working hypothesis is that a fresh institution or agency B is better placed to carry out a particular function than the old regime A. Specific responsibilities are handed over rather than the entire care process (which is the domain of level 1 change). Day-to-day improvements are deemed to flow by dint of better specialist management – the new body produces a better division of labour, better cost containment, better information flow, better training, and so on.

Referral management centres (RMCs) or gateways (RMGs) are the perfect exemplification of this mode of change. A new body is transplanted into the midst of the care pathway to assume responsibility for core referral decisions. Much ink has been spilt describing the potential of these agencies in accompanying plans, proposals and position papers. In these narratives RMCs are envisaged as strengthening organisational capacity by ensuring that referrals are clinically necessary and that patients see the right person, in the right place at the right time. For example, a position paper from a regional Medical Committee explains that "Not all

referrals require a consultant opinion” and that “Clinical assessment of referrals should mean that the most appropriate person sees the patient – quickly, close to home, and at the lowest reasonable cost” (Somerset Local Medical Committee, 2004, p1).

These particular interventions and their underlying theory have experienced what can only be described as a torrid history. General Practitioners (GPs) have been particularly vociferous in expressing concerns about possible negative consequences of RMCs. A reverberative counter theory is that RMC staff are less able to review referrals accurately because they lack accumulated wisdom on the particular patient or because they lack the clinical skills, knowledge, expertise or experience (Collins, 2012; Nowotny, 2011). However, if we allow history to take its course, RMC theory comes in for gradual modification in which these counter theories are countered, as in the following excerpt on the ‘Manchester Gateway’ (Wootton and Whiting, 2012, p3):

‘With the exception of one or two practices, the vast majority of GPs from the 103 practices now use the gateway ... a number of factors have led to this almost universal acceptance. The first is that this was not something introduced suddenly and on a whim. It was developed over a long period of time and after plenty of discussion and collaboration with primary care through local meetings ... One very important factor in setting up the gateway was the use of ... senior, well-known GPs. It meant the other GPs respected the decisions made and the feedback they were receiving. Because the advice is provided in a very professional and supported way, the GPs saw it as more of a [professional development] tool than criticism of their decisions’

Once again we refrain from questioning the veracity of these claims. Our point is that the underlying theory has transmuted. Wholesale management takeovers are controversial. What starts as a level 2 theory about superior administrative capacity, also turns out to depend on role and procedural adaptations (level 3 and 4) in order to calm the protestations of the old regime.

#### 4.3. Role change

The third model proposes less dramatic change in which systems and organisational structures remain intact but within which roles and responsibilities are shifted. Decision making is transferred into the hands of those who are considered better placed to interpret and act upon referral options. One such example is providing GPs with direct access to diagnostic tests. The idea is to allow GPs to bypass the gatekeeper role of the consultant in accessing tests results (Sibbald, 2009). The assumption is that, armed with immediate data, GPs will manage some patients directly within primary care thus reducing the demand on the secondary system. Essentially under this role reallocation model, GPs take on minor aspects of the consultant’s brief.

Critics express concerns that GPs may then over-utilise diagnostic tests and also to question whether GPs have the knowledge and skills to interpret the tests or to take the optimal action on the basis of the results (Robinson, 2011). For example, a 2008 *Pulse* article features an interview with a GP with a special interest in musculoskeletal conditions who explains that a direct access scheme for MRI to investigate back or shoulder problems did not work because “GPs had not been comfortable in interpreting the results” (Nowotny, 2008, p1).

Consequently and cautiously, direct access provision has been supplemented with other interventions to ensure GPs access and interpret the tests appropriately. Supplementary manoeuvres include administrative (level 2) change in the form of triage of GP

referrals or the addition of ‘all in one’ diagnostic and interpretation services’ into the patient pathway. For example, the DH website (<http://healthandcare.dh.gov.uk/direct-access-diagnostic-tests/>, accessed 26/09/13) features one example of a private provider which offers an all-in-one diagnostic and interpretation service for ECG test results so that “ECG results can be interpreted immediately thereby having great potential to save NHS money by reducing the need for outpatient referrals and improving patient care by enhancing patient convenience and reducing waiting times for diagnostic tests”.

This model has also begun to incorporate procedural (level 4) changes in the form of guidelines or feedback to encourage GPs to review and change their use of diagnostic tests. For example, in 2012, the DH produced a set of guidelines for GPs which set out the circumstances in which patients with symptoms of four different cancers (ovarian, lung, colorectal and brain) should be referred directly for diagnostic tests (Department of Health (2012)). DH is also working “to ensure that data are routinely collected about GP usage of these tests, so that GPs can benchmark their use of them” ([http://www.ncin.org.uk/collecting\\_and\\_using\\_data/data\\_collection/diagnostic\\_tests.aspx](http://www.ncin.org.uk/collecting_and_using_data/data_collection/diagnostic_tests.aspx), accessed 26/9/2013). Here again, the core theory about how role change is supposed to work has morphed to incorporate the other stanchions of organisational change.

#### 4.4. Procedural change

The fourth motor of change is at the level of individual practice. It requires minimal system reorganisation; all basic pathways, organisations and roles remain in place. A typical example of this particular programme theory is the provision of guidance or feedback to allow those making referrals to reflect upon and reshape their own decision making. One example of this is the use of peer review of referral decisions, which utilises pooled knowledge and expertise to give advice on the appropriateness of individual referral decisions (Chambers, 2010). The idea is that reviewing cases in a supportive and educational manner allows for the assimilation of best practice over the course of time, which ultimately normalises referral patterns between individual GPs and GP practices (Maggs-Rapport et al., 1998; Wilczynski, 2012).

Counter theories suggest that there can be a net gain in knowledge from peer review but with no ultimate change in referral behaviour, particularly if the feedback is not timely. For example, a 2012 *Pulse* (p1) article discusses the initial set up of a peer review scheme in Manchester that had little effect on referral rates “while we identified learning opportunities, it seemed to do little to actually change referral behaviour. The main problem was that the data was always two to three months out of date” (Wootton and Whiting, 2012). This experience was mirrored in a review of a peer review scheme from a local CCG (Coastal West Sussex Clinical Commissioning Group, 2011, p2). “The group believes the reduction opportunities will be realised more robustly if the process involves a prospective rather than retrospective review, as this challenges behaviour in ‘real time’ and therefore helps embed the learning”.

With this leap in thinking peer review becomes a form of triage (level 2 change) rather than feedback and, consequently, other structures are needed to allow this to happen in a timely manner. The solution for the Manchester CCGs was to have senior GPs triaging referrals and a central booking scheme managing data flow (Wootton and Whiting, 2012). Again, the effectiveness of the procedural (level 4) changes becomes dependent on the implementation of organisational (level 2) change. Furthermore, 2011 saw the inclusion of indicators relating to peer review of referrals into the Quality and Outcomes Framework (QOF) which rewards GPs in the UK for implementing good practice. QOF7 rewards



practices that participate in external peer review to compare secondary care outpatient data. Procedural change through peer review becomes incentivised, suggesting that level 5 support may be needed to secure procedural change.

#### 4.5. Motivational change

The final theory of change (level 5) includes the practice of offering incentives to encourage GPs to reshape their referral behaviour. Change is located squarely in the hands of the individual practitioner; all system pathways, organisations, roles and routines remain in place. Improvements are deemed to follow from confronting individuals with financial carrot or stick to change and/or limit their referral practices. The intended change is behavioural and driven by personal interest rather than peer learning.

As noted above, indicators relating to the peer review of referrals and the development of care pathways were added to QOF indicators in 2011. In addition, some CCGs/PCTs have implemented 'referral incentive schemes' whereby GP practices receive payments for meeting targets to cut referral rates. A recent *Pulse* article entitled 'Five questions you should ask before offering referral incentives' suggested "Under a well-planned scheme, commissioners have the possibility of improving outcomes and to make care more cost effective, releasing money to be spent on other services" (Poole, 2012).

Counter theories postulate that referral incentives can encourage a blanket approach to all patients with a similar condition thus failing to take account of individual patient differences and their referral needs. In this way they may introduce "unacceptable conflict into the doctor patient relationship" particularly if decisions are felt to be motivated by cost rather than clinical evidence (Poole, 2012). Indeed, the proliferation of referral incentive schemes that match rewards to specific referral targets and therefore appear to reward arbitrary reductions in referral activity have been widely condemned by the British Medical Association (BMA, 2013). Significant concerns are expressed regarding the professional and ethical implications of target driven referral incentive schemes where there is no concurrent assessment of the appropriateness of such reductions: "It is not acceptable for practices to continue reducing their referrals to secondary care when this means that decisions are being taken that are not clinically appropriate and will have a detrimental effect on the healthcare of patients." (BMA, 2009, p3).

In order to be successful, revised programme theories advocate that incentives need to be targeted upon outcomes that have a consensus regarding clinical importance (Roland, 2004). As a *BMJ* article opines, GPs respond positively to incentives for peer review of referrals because "it's a very good way of learning from one another" (Eaton, 2008). Here we see that motivational stimuli (level 5) appear to require the support of some mode of interpersonal adaptation (level 4) in order to maintain the delicate balance of quality and volume of referrals.

Our analysis above reveals a typical feature of the programme architecture, that there is never just the single blueprint. Plans change. Programme designers adapt their ideas in the face of criticism and by dint of their own and others' experiences. Our analysis of programme theories rarely supplied an agreed model but revealed a history of planning. In terms of long running interventions, like DM/RM, that history tends to build a picture of the threats that should be taken into account in the delivery of successful programmes. The identity of these necessary conditions corresponds closely to the analysis made earlier under the notion of the depth ontology of organisational change. These contingencies, these other things that one must also modify to make change happen at one level are immediately identifiable in the remaining layers of the structure.

Thus, we find that the framework depicted in Fig. 1 has a deleterious consequence on the evidence that ensues. Without doubt, significant change may be provoked at any of these levels. However, regardless of where it begins, the progress of any such intervention will also involve a considerable journey as all the other layers absorb the change in interlocking and temporally complex ways. In our analysis above, we show that what is provoked at one level may be prevented at another. Accordingly, researching service improvement really requires an evaluation of the flows and blockages of any particular programme theory around the entire system. A basic illustration of the strategy is provided in Fig. 2.

## 5. Organisational change as encountered in research review

In this section we question the extent to which existing reviews of referral management interventions are equipped to take account of our revised theory depicted in Fig. 2. From its origins in clinical research, systematic review began with the idea of examining *all* studies on *one* type of intervention. Increasingly, this situation where the reviewer examines *many* studies of *several* different interventions is becoming the norm.

Confronted with this level of complexity a new model of research synthesis is emerging that might be termed a 'typological review'. The basic format is summarised in the matrix in Fig. 3. Firstly a decision is made on the significant types and sub-types of the broad intervention under review (types A to E in the diagram). Most complex interventions have multiple ambitions and the second element of the review structure is to decide on the various outputs and outcomes in terms of which the different family members will be assessed and compared. The schema is completed in the cells of the matrix where the findings are arrayed. The review calls on a variety of evidence to provide a 'scorecard' or 'dashboard summary' of each species in terms of whether the evidence is broadly positive (✓), broadly negative (×) or absent/indecisive (?). The overall objective of the typological review is to distinguish the most effective subtypes (depicted notionally in the diagram as type C). However, by using this approach the evidence-base itself becomes stratified and does not take account of temporal relationships between different organisational strata.

We have identified five previous DM/RM reviews (Akbari et al., 2008; Imison and Naylor, 2010; Jack and Powers, 2009; Martin et al., (2010); Roland et al., 2006). We concentrate on the common challenges they face, most particularly in respect of our core thesis about the constantly shifting ground of intervention theory.

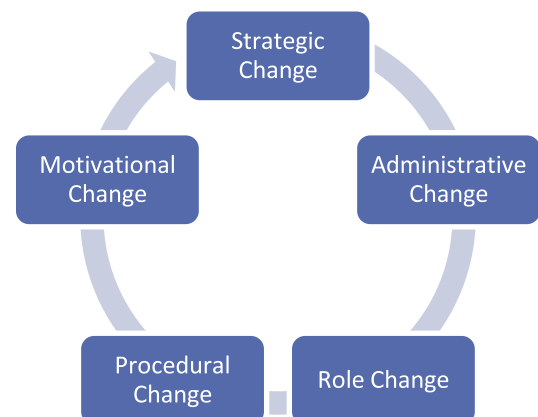


Fig. 2. Interconnected levels of organisational change.

Intervention	Output 1	Output 2	Output 3	Outcome 4	Outcome 5
Type A	✓	✗	✗	?	✗
Type B	✗	✓	✗	✗	?
Type C	✓	?	✓	✓	✓
Type D	✗	✓	✓	✗	?
Type E	✓	✗	✓	?	?

Fig. 3. The typological review.

If our previous analysis is correct it would suggest that the typological review, because of the very way it is constructed, will face a number of severe methodological challenges. Demand management interventions have a natural tendency to hybridise. In order to work: strategic innovations will need to be buttressed over time with administrative reforms; changes in everyday duties are more likely to bed down if supported by role realignment and a motivated workforce ... and so on. This suggests formidable challenges ahead for any attempt to lay down a classification system in the expectation that a review will then go on to encounter pure empirical instances of each type. In particular, we might anticipate stumbling blocks in terms of: i) disagreement in the definitions of the operational types and sub-types, and ii) indeterminacy in being able to assess and differentiate clear outcomes corresponding to each category.

We explore the first of these conjectures in Table 1, which summarises the different schemata employed in the five reviews. It reveals both promising similarities and unsettling differences between the schemas. Many of the same categories and sub-categories are discernible. There is, however, no consensus on how to subdivide the field. The coverage, number, naming and content of each type and sub-type differs from review to review. In terms of coverage, Jack and Powers (2009) stretch the field massively to cover demand, capacity and performance management schemes, whereas Roland et al. (2006) review only those

schemes involved in establishing specialist outpatient services within the primary care environment. Even when there is overlap in the broad categories, the 'same' interventions are classed differently. For example, 'in house second opinion by a GP' is conceptualised as a 'professional behaviour change' intervention by Roland et al. (2006), an 'organisational intervention' by Akbari et al. (2008), a 'managerial intervention' by Martin et al., (2010) and a 'clinical triage and assessment' intervention by Imison and Naylor (2010).

One method for improving discernment in a typology is to identify sub-types within each domain. Such sub-categorisation is achieved by moving down a level of abstraction. Rather than generic titles like 'organisational' or 'behavioural' interventions, the sub-schemes are then identified using practitioner or domain-specific NHS terminology (e.g. 'consultation liaison psychiatry', 'PCT-led referral management systems'). The Akbari et al. (2008) review employs this strategy in the extreme with the subtypes or sub-subtypes being represented by individual or very small batches of interventions (small, incidentally, because in this the interventions need to have been researched in primary studies that meet Cochrane methodological requirements).

Does the fragmentation of a classification system resolve the problem of the intermixing of intervention theory? We think not, for the very idea of typological review is to serve evidence based management by differentiating approaches and choosing between

**Table 1**  
Types identified in demand management reviews.

Roland et al. (2006)	Jack and Powers (2009)	Akbari et al. (2008)	Martin et al., (2010)	Imison and Naylor (2010)
<b>Transfer of services to primary care</b> (surgical/ medical clinics in primary care, GPSIs, changing outpatient follow up, GP direct access to diagnostic tests or services) <b>Relocation of services to primary care</b> (outpatient clinics held in primary care, telemedicine, attachment of specialists to primary care teams) <b>Joint working between primary and secondary care</b> (shared care management of chronic disease, consultation liaison psychiatry) <b>Professional behaviour change</b> (structured referral sheets, educational outreach by specialists, in house second opinion by a general practitioner, financial incentives)	<b>Demand modelling</b> <b>Health Maintenance Organisations</b> <b>Vertical/Horizontal Integration</b> <b>Capacity management</b> <b>Workforce management</b> <b>Resource Utilization</b> <b>Subcontracting</b> <b>Information Technology</b> <b>Quality of Care</b> <b>Technical efficiency</b> <b>Financial incentives</b>	<b>Professional education interventions</b> (passive dissemination of local guidelines, guidelines with structured referral sheets, secondary care provider led educational strategies) <b>Organisational interventions</b> (attachment of specialists to primary care teams, in-house second opinion, changes to appointment systems based on size of practice) <b>Financial interventions</b> (change in remuneration system, GP fundholding, change in patient charging systems)	<b>Professional education interventions</b> (structured referral sheets, educational activities by secondary care providers, passive dissemination of referral guidelines) <b>Managerial interventions</b> (in-house second opinion, primary care clinics for chronic disease, changing outpatient follow up, GP direct access to diagnostic test, GP direct access to services, GPs with a Special Interest, Referral Management Schemes) <b>Financial interventions</b> (incentives to GPs)	<b>Referral Management Centres</b> <b>Clinical Triage and Assessment</b> (eg, in-house second opinion, community based specialist clinics, specialists attached to primary care teams, phototriage) <b>Peer Review and feedback</b> (eg, peer review of referral letters, educational outreach visits by specialists) <b>Financial Incentives</b> (eg use of referral standards in contracts, PCTs led referral management system) <b>Guidelines plus other support</b> (eg structured referral sheets) <b>Passive use of guidelines</b>

them. The more the sub-type identified is specific to local implementation and contextual conditions then the longer become the odds of reproducing it elsewhere. It becomes much more difficult to identify a transferrable formula when the methodological entreaty is to adhere to ‘type 4, subtype 5’. For instance, one successful subtype turns out to be a multi-faceted intervention involving educational meetings, a new referral and reply sheet, new staff and changes in equipment and facilities (Thomas et al., 2003). What type or sub-type is under consideration here? How could one follow that template? The paradox remains: typologies are simplifications and accentuations – interventions are complex and adaptive.

We turn now to our second hypothesis in relation to the findings of the typological review. If programme typologies fail to recognise interdependencies between programme theories, if they mask the depth ontology of organisation change, what is the fate of the outcome ‘scorecards’? Somewhat against our contention that they will end in indeterminacy, we begin with instances from our set of reviews that claim to identify winners and losers. In the latter category, the most persistent finding from the reviews is that the ‘passive use of guidelines’ is ineffective in controlling demand through changing GPs referral behaviour.

The explanation is that if referral guidelines are simply issued to GPs they have little support or incentive to follow them – given the scores of other demands on their time. However, this explanation is a perfect exemplification of our overall thesis that successful innovations depend on action up and down the organisations strata. To quote Akbari et al. (2008) (p11) guidelines may be effective if: ‘local secondary care providers are involved in dissemination activities, structured referral sheets are used, secondary care management is responsive to changes in primary care behaviour as a result of the guidelines and if they reflect local circumstances and address local barriers’. In the right conditions the all-time loser can become a winner.

With this possible exception, our claim is that typological scorecards tend to the indeterminate. Firm conclusions are limited by study selection. DM/RM interventions are colossal in number but are represented in Akbari et al.’s (2008) review by 17 studies and in Roland et al.’s review (2006) by 119 inquiries. The former uses the customary Cochrane restriction to primary studies conducted by RCTs and a fair proportion of its discussion is given over to the further methodological weaknesses of the subset that has passed initial muster. This is the entirely typical root of ‘indeterminacy mark I’. Reviews proceed rigorously forward but by passing down responsibility to poorly executed primary studies it becomes ‘difficult to draw firm conclusions’.

The second root of indeterminacy is the proper caution expressed by authors as a result of the mixed outcomes that typically emerge in typological reviews. To appreciate this we examine a real scorecard (Table 2) extracted from analysis (Sibbald et al., 2007), derived from Roland et al.’s (2006) review, and compare it with the idealised framework with its notional ‘winners’ and ‘losers’ as in Fig. 3. This table covers three different demand management schemes involving the relocation of services to primary care. First, one notes the equivocation expressed in many of the cells. There are partial gains here and losses there but the majority of the cells report ‘insufficient evidence’, ‘no change’, ‘variable outcomes’. Also noticeable but hard to quantify is a counterbalancing tendency whereby a gain on one outcome is met by loss in another quarter.

There are a couple of reasons for this mixed picture. ‘Indeterminacy mark II’ is rooted in the classic problems of under-reporting and non-equivalence of analytic frameworks in primary studies. In total Roland’s team seek to review four major intervention modes, broken down into 15 subtypes, analysed on 6 different outcomes

**Table 2**  
Summary of findings from Sibbald et al. (2007) on the relocation of services to primary care.

Intervention	Outcomes					
	Access/equity	Quality/health	Hospital impact	General practice impact	Health service cost	Implementation issues
Shifted outpatient clinic	Improved access: potential to improve equity if located to populations with poor access to secondary care	Insufficient evidence: in theory, quality should be unchanged	Insufficient evidence on outpatient use: some patients will require added outpatient visit because primary care lacks diagnostic facilities	Insufficient evidence on workload: no gains in GP knowledge or skills	Clinics serving urban advantaged populations are not cost-effective due to loss of economies of scale	Requires expansion in specialist workforce to compensate for loss of economies of scale
Telemedicine	Improved access for remote populations; potential to improve equity of located to populations with poor access to secondary care	Insufficient evidence on health outcomes; diagnosis more difficult for some specialities (ie dermatology) but may improve with advances in technology	Insufficient evidence on outpatient use: some patients will require added outpatient visit because primary care lacks diagnostic facilities	Insufficient evidence but likely to increase primary care workload	Cost effectiveness is highly context dependent but generally better when telemedicine clinics are located in remote areas where patient travel costs to outpatient clinics are high	Requires substantial investment in equipment and training of clinicians
Attachment of specialist	Improved access	No change	Variable: reduced outpatient referrals in some specialities (physiotherapy) but not others	No change	Variable: appears cost effective for some specialities but not others	Requires expansion if specialist workforce and deployment to primary care teams

(‘equity’, ‘quality’, ‘hospital impact’, ‘GP impact’, ‘cost’ and ‘implementation issues’). Typological high ambition is met with paucity in the coverage and content of primary studies. There is no reason why a one-off evaluation should employ the generic framework, resulting in inevitable absence of information in some cells.

This brings us to the all-pervasive ‘indeterminacy mark III’. If one examines closely the summary findings deposited in most cell entries they are not in fact verdicts but, quite properly, conditional statements. To use an obvious illustration, the first ‘results cell’ in Table 2 (Sibbald et al., 2007) tells us that relocating an outpatient clinic will improve access and equity *if* it is located in the appropriate (badly-served) area. In another example it transpires tele-medicine is more likely to be cost efficient *if* utilised in remote areas. What is being discovered, unsurprisingly, is that there are no universal panaceas and that conditionality is the norm. This feature is most evident in the final column of the Sibbald et al. (2007) evidence matrix, where sits a brief assessment of ‘implementation issues’. In each and everyone one there is a ‘condition’ that for the sub-scheme to work, other resources need to be brought to bear and other working practices must be changed to fall in line with the developing programme ... scheme X requires expansion of a specialist workforce, scheme Y depends on good communication between actors, scheme Z requires major revision in working practices and so on.

Inexorably, we are returned to our core thesis that, whatever the initial orientation of a demand management scheme, for it to work requires the coordination of strategies, rules, roles, routines and individual support. Alas, policy makers have always displayed considerable antagonism when served up contingent findings in evaluative inquiry and research reviews. And this distaste results in a final characteristic of the scorecard summaries of the typology review, namely their free use of the model auxiliary verb. Of the various approaches it is frequently concluded that ‘it *may* improve the referral process’, ‘it *may* not improve the quality of referrals’, ‘it *appears* a promising approach’, ‘it represents a *plausible* strategy’, ‘it is *generally* attractive’. Because they cannot fulfil the grand ambition of sorting intervention wheat from chaff, scorecards have to retreat from probabilities to possibilities. We suppose that ‘possibilities’ are a poor guide to policy making. The proper answer to the ‘what works’ question is always ‘it depends’ and the job of research synthesis, we submit, is to deepen our understanding of the contingencies.

## 6. Conclusion

We conclude with the answer to the question posed in our title. It turns out that reviews of healthcare programmes often miss lessons vital in understanding how to improve healthcare systems. Lasting system transformation depends on the dovetailing of strategic change, administrative change, role change, procedural change and motivational change. Rarely does evaluation research or systematic review address this wider scenario.

So how might such a depth of vision be integrated in research synthesis? Elsewhere and at book length we have attempted to provide a range of answers to this question (Pawson, 2013). Here we make a paragraph length plea for just one promising line of inquiry, namely that both evaluation and systematic review should study programme history. Evaluation research forever chases newly minted programmes and primary research of this ilk also provides the bread and butter of evidence for systematic review. More attention should be paid to longstanding programmes because they stand longest, having dealt with the system strains described above. The miniature programme histories outlined above have the same structure. What they all say is that in order to

achieve A, we had in mind intervention B, but to make it work we had to deal with contingencies D, E, and F. These reflections are practical programme theories. The evidence to test and refine them is already out there.

## Funding

‘Motivations, mandates, accountability and patient choice: a realist synthesis of demand management’ is funded by the UK National Institute for Health Research Health Services and Delivery Research Programme 11/1022/04. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HS&DR Programme, NIHR, NHS or the Department of Health.

## Acknowledgements

We would also like to acknowledge additional contributions to the project from: Roberto Longo, Lisa Maginnis, Steve Montague, Hugh Sturgess, Laurence Wood, Judy Wright.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.socscimed.2014.05.032>.

## Web pages

[http://www.ncin.org.uk/collecting\\_and\\_using\\_data/data\\_collection/diagnostic\\_tests.aspx](http://www.ncin.org.uk/collecting_and_using_data/data_collection/diagnostic_tests.aspx)  
<http://healthandcare.dh.gov.uk/direct-access-diagnostic-tests/>

## References

- Akbari, A., Mayhew, A., Al-Alawi, M., Grimshaw, J., Winkens, R., Glidewell, E., et al., 2008. Interventions to improve outpatient referrals from primary to secondary care. *Cochrane Database Syst.Rev.* 8 (4), CD005471 <http://dx.doi.org/10.1002/14651858.CD005471>.
- Archer, M., 1995. *Realist Social Theory: the Morphogenetic Approach*. Cambridge University Press, Cambridge.
- Archer, M., 2003. *Structure, Agency and the Internal Conversation*. Cambridge University Press, Cambridge.
- Best, A., Greenhalgh, T., Lewis, S., Saul, J.E., Carroll, S., Bitz, J., 2012. Large-system transformation in health care: a realist review. *Millbank Q.* 90, 421–456.
- British Medical Association, 2009. *GP Referral Incentive Schemes: GPC Guidance for GPs and LMCs*.
- British Medical Association, 2013. *Conflicts of Interest in the New Commissioning System: Doctors as Providers*.
- Byrne, D., Callaghan, G., 2013. *Complexity Theory and the Social Sciences: the State of the Art*. Routledge, London.
- Chambers, R., 2nd December, 2010. *How to Peer Review Practice Referrals*. GP Online.
- Clark, A., 2013. What are the component of complex interventions? Theorizing approaches to parts, powers and the whole intervention. *Soc. Sci. Med.* 93, 185–193.
- Coastal West Sussex Clinical Commissioning Group, 2011. *Referral Management Peer Review Scheme – 6 Month Review*.
- Collins, G., 2012. Almost Half of GPs Say Delays Caused by Referral Management Schemes have Harmed Patients. *Pulse*, 27th September.
- Department of Health, 2012. *Direct Access to Diagnostic Tests for Cancer: Best Practice Referral Pathways for General Practitioners*. Department of Health, London.
- Eaton, L., 2008. BMA backs GPs in their objections to financial incentives to limit hospital referrals. *Br. Med. J.* 337, a2306.
- Erlang, A.K., 1909. The theory of probabilities and telephone conversations. *Nyt. Tidsskrift Matematik B* 20.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O., 2004. Diffusion of innovations in service organizations: systematic review and recommendations. *Millbank Q.* 82, 581–629.
- Hage, J., Aiken, M., 1970. *Social Change in Complex Organizations*. Random House, New York.
- Hendry, C., 1996. Understanding and creating whole organizational change through learning theory. *Hum. Relations* 49, 621–641.
- Imison, C., Naylor, C., 2010. *Referral Management: Lessons for Success*. The Kings Fund, London.



- Jack, E.P., Powers, T.L., 2009. A review and synthesis of demand management, capacity management and performance in health services. *Int. J. Manag. Rev.* 11, 149–174.
- Jones, D., Mitchell, A., 2006. *Lean Thinking for the NHS: a Report Commissioned by the NHS Confederation*. The NHS Confederation, London.
- Maggs-Rapport, F., Kinnersley, P., Owen, P., 1998. In-house referral: changing general practitioners' roles in the referral of patients to secondary care. *Soc. Sci. Med.* 46, 131–136.
- Martin, A., Macleod, C., Naqvi, S.A.R., 2010. Effectiveness and cost effectiveness of targeted interventions to reduce unnecessary referrals and improve the quality of referrals from primary care to secondary care: a rapid review of the research literature. Evidence Adoption Centre East of England, Cambridge.
- Merton, R., 1967. *On Theoretical Sociology: Five Essays Old and New*. Free Press, New York.
- Murray, E., Treweek, S., Pope, C., MacFarlane, A., Ballini, L., Dowrick, C., et al., 2010. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. *BMC Med.* 8, 63.
- NHS Modernisation Agency, 2002a. In: D.O. Health (Ed.), *The Big Wizard*. Department of Health, London.
- NHS Modernisation Agency, 2002b. *The Little Wizard*. Department of Health, London.
- NHS Scotland, 2007a. *Patient Flow in Planned Care: Admission, Discharge, Length of stay and Follow-up*. The Scottish Government, Edinburgh.
- NHS Scotland, 2007b. *Patient Pathway Management: Referral Facilitation*. The Scottish Government, Edinburgh.
- Nowotny, S., 2008. MRI Study Boosts Drive for Direct GP Access to Diagnostic Tests. *Pulse* (p. 10th November).
- Nowotny, S., 2011. Gatekeeper Job is One for Trained GP. *Pulse* (10th August).
- Pawson, R., 2006. *Evidence Based Policy: A Realist Perspective*. Sage, London.
- Pawson, R., 2013. *The Science of Evaluation: a Realist Manifesto*. Sage, London.
- Pettigrew, A., McKee, M., Ferlie, E., 1988. Understanding change in the NHS. *Public Admin.* 66, 297–317.
- Poole, S., 2012. Five Questions You Should Ask before Offering Referral Incentives. *Pulse* (4th October).
- Robinson, S., 2011. Exclusive: PCTs Restrict Access to Diagnostic Scans (GP Online, 8th September).
- Roland, M., 2004. Linking physicians' pay to the quality of care – a major experiment in the United Kingdom. *N. Engl. J. Med.* 351, 1448–1454.
- Roland, M., McDonald, R., Sibbald, B., 2006. Outpatient Services and Primary Care: a Scoping Review of Research into Strategies for Improving Outpatient Effectiveness and Efficiency.
- Sibbald, B., 2009. Direct access to diagnostic services. *Br. J. Gen. Pract.* 59, 562.
- Sibbald, B., McDonald, R., Roland, M., 2007. Shifting care from hospitals to the community: a review of the evidence on quality and efficiency. *J. Health Serv. Res. Policy* 12, 110–117.
- Somerset Local Medical Committee, 2004. *LMC Position Paper on Referral Management Centres*. Somerset Local Medical Committee.
- Thomas, R., Grimshaw, J., McClinton, S., 2003. Cluster randomised controlled trial of a guideline based open access urological investigation service. *Fam. Pract.* 20, 646–654.
- Thomson, S., Osborn, R., Squires, D., Jun, M., 2012. *International Profiles of Health Care Systems, 2012*. The Commonwealth Fund, New York.
- Wilczynski, P., 2012. Taking the Isolation out of Referral Decisions. *Pulse* (3rd September).
- Willmott, R., 2000. The place of culture in organization theory: introducing the morphogenetic approach. *Organization* 7, 95–128.
- Wootton, S., Whiting, M., 2012. Creating Ownership of a Referral Gateway. *Pulse* (3 September).
- Zaltman, G., Duncan, R., Holbeck, J., 1973. *Innovations and Organisations*. Wiley, New York.