We need better descriptions than ‘evidence-based policy’ and ‘policy-based evidence’: the case of UK government ‘troubled families’ policy

Abstract. ‘Evidence-based policymaking’ and ‘policy based evidence’ are political slogans rather than good descriptions of the use of evidence in policy and policymaking. To produce more meaningful categories we need to establish clear criteria based on the combination of evidence, values, and judgement in political settings. To that end, I use policy theories to establish the limits to the use of evidence in policy, and the case study of ‘troubled families’ policy to show how governments use evidence. This case study is useful because, while it is easy to conclude that the UK Government has broken all of the rules of evidence-based policymaking, it is not easy to describe a politically feasible alternative.

Introduction

‘Evidence-based policymaking’ (EBPM) is a poor phrase because no one really knows what it means. One can never be sure if a policy was ‘evidence-based’ because there are no clearly defined and agreed upon criteria or principles on which to draw. Rather, it is used primarily as a political slogan, by scientists to support a far greater role for scientific evidence in policy, or by policymakers to argue that a government is already taking scientific evidence into account and funding well-evidenced programmes (Cairney, 2016a).

Its political comparator is ‘policy-based evidence’ (PBE). As an accurate description of policymaking, PBE has similar problems to EBPM. It is often invoked when critics, disagreeing with specific government policies, argue that a policymaker decided what they wanted to do then sought information to back it up. Yet, there are many other practices included in this umbrella term, including: responding to the scientific evidence of a problem but combining ideological, economic, and scientific factors to choose a solution; inflating the evidence of a solution’s effectiveness to make a decisive choice quickly; and/ or using the wrong kinds of knowledge to identify problems and evaluate and deliver policy solutions (Cairney, 2016b). Further, critics often describe PBE when policymaking in the real world does not live up to an unrealistic EBPM standard (Oliver et al, 2014a; 2014b). They identify the insufficient use of evidence during an artificially singular moment of choice - the policy - when, in fact, the policy process, and use of evidence, is continuous. Or, they fail to consider how the need to act quickly despite uncertainty affects the standards we set on evidence gathering, and informs the balance between evidence and political judgement (Cairney et al, 2016a). Therefore, it is difficult to distinguish between egregious PBE and more defendable practices.

My aim is to go beyond these political slogans – and build on more realistic discussions of evidence informed policy (Nutley et al, 2007) - to highlight the general limits to the use of evidence in policymaking. I identify categories to help us generate more thoughtful evaluations when elected policymakers pursue many kinds of ‘good’ decisions. They need to maintain electoral popularity and make decisions more quickly than we like, prompting: pragmatic
choices about the availability of scientific evidence, value judgements to prioritize problems and target populations, and negotiation with other actors to make choices and ensure delivery.

First, I outline key insights from policy theories to clarify the meaning of EBPM in the real world. I identify the extent to which policymakers have to use major shortcuts to gather evidence and make decisions quickly in a complex policymaking system, the dilemmas involved in combining evidence-based policies with principles-based governance, and the use of different criteria and methods to combine evidence with judgement (Dowding, 2015). It helps us distinguish between many categories of alleged PBE.

Second, I use the case study of the UK Government’s ‘troubled families’ policy because it appears to be an extreme case: initial support for ‘family intervention’ approaches represented a ‘classic case of policy based evidence’ (Gregg, 2010), and policymakers cherry picked evidence to justify massive expansion and defend a top-down, evidence-light, and quick emotional reaction to crisis (Crossley, 2015a). Within one programme are many kinds of problematic evidence use, including estimates of the problem, measures of success, and interpretations of academic evaluations and neuroscience. This discussion allows us to identify different uses of evidence within a single programme, compare types of PBE, and situate them within the wider context of Westminster politics. I argue that the order of comparison influences the ways in which we characterise PBE: one story highlights the overarching role of egregious PBE to declare government success; another highlights a perception in government that the ends (good policy) can justify the means (bad evidence).

**The meaning of EBPM in the real world: don’t confuse it with EBM**

It is difficult to say what will satisfy critics of PBE if they compare policymaking to an undefined standard. ‘Evidence based medicine’ (EBM) is a useful starting point because its ideal is influential across several disciplines - including health, environmental and management sciences – but produces a naïve expectation for EBPM (Oliver et al, 2014b; Cairney, 2016a: 65; Cairney et al, 2016; Morrell, 2008). Despite more modest intentions (Sackett et al, 1996), EBM has become associated with an agenda to: (a) gather the best evidence on interventions in a central hub, based on a hierarchy of methods, in which randomised control trials (RCTs) and their systematic review are at the top; and, (b) ensure that it has a quick and direct impact on practice (Oliver et al, 2014b; Cairney et al, 2016a).

This approach contrasts with ways in which policymakers use evidence: there is no equivalent to a hierarchy of evidence, and the policymaking environment involves a larger, more heterogeneous, set of actors, producing an indirect relationship between evidence and action (Lomas and Brown, 2009: 906; Elliott and Popay, 2000: 467; Stoker, 2010: 53; Bédard and Ouimet, 2012: 625). Consequently, evaluations of EBPM referring to EBM rather than policy theory (Embrett and Randall, 2014) do not fully appreciate three ever-present factors which produce the appearance of PBE in all policy choices.

1. **Policymakers have to take major short cuts to gather information quickly**
People use two shortcuts to gather information quickly: ‘rational’ ways to establish the best available evidence and sources of evidence, and ‘irrational’ ways to understand policy problems, drawing on emotions, habits, and deeply-held beliefs (Kahneman, 2012: 20; Haidt, 2001: 818; 2007; 2012; Alter and Oppenheimer, 2009: 220). The accumulation of scientific knowledge, and large capacity of government, does not solve this problem (Simon, 1976; Botterill and Hindmoor, 2012; Lewis, 2013: 9-10; Schneider and Ingram, 1993; Schneider et al, 2014; Pierce et al, 2014). Policymakers have too many problems to pay attention to, too many solutions to consider, and too many choices to make, based on more information than they can process. So, they combine judgements based on well-established beliefs, and informational shortcuts based on familiarity with information, even if they are committed to evidence-informed processes.

In such circumstances, ‘the evidence’ seems secondary to the ways in which policymakers understand it. They are receptive to particular kinds of evidence – to address the problems to which they pay most attention, and provide solutions consistent with their beliefs or knowledge – and ways in which the evidence is ‘framed’, such as to appeal to the emotions and the familiar (Dearing and Rogers, 1996: 1; Baumgartner and Jones, 1993: 11-2; Kingdon, 1984: 3–4; Cairney, 2012: 183). Policy theories identify the links between evidence and persuasion when policy actors: combine facts with emotional appeals, to prompt lurches of attention (True et al, 2007: 161); produce feasible policy solutions and exploit times when policymakers have the motive to adopt them (Kingdon, 1984); tell stories which manipulate biases, apportion praise and blame, and highlight the moral value of solutions (Jones et al, 2014); and, interpret new evidence through the lens of established beliefs (Weible et al, 2012).

2. They use evidence in an environment or system over which they have limited control

A direct link between evidence and action requires something akin to a singular moment of authoritative choice - the policy – made and implemented in a ‘policy cycle’ with key ‘stages’ including agenda setting, policy formulation, legitimation, implementation, and policy maintenance, succession, or termination (Hogwood and Gunn, 1984). Yet, there are two key problems with a stage-based understanding (Cairney, 2016a: 18; 2012: 34; 2014; John, 2012; Sabatier, 2007; Everett, 2003; Colebatch, 2006). First, it implies a core group of policymakers and analysts at the heart of the process, making policy from the top down, without recognising the diffusion of policy responsibilities across multiple venues. Second, it downplays the tendency of policy to be made continuously as it is delivered, or as decisions to solve one problem intersect with decisions on many related policies. Instead, most policy theories identify multi-level policymaking environments with five key characteristics:

- A wide range of actors (individuals and organisations) seek to make or influence policy at many levels of government and in multiple venues.
- A proliferation of rules (‘institutions’) inform behaviour in each venue. Some rules are written down and easy to follow. Others are implicit and only visible to actors with the time to engage.
• Policymakers and influential actors develop networks built partly on trust and the regular exchange of information. Networks often develop out of the public spotlight and at a low level of government.

• Certain ways of thinking dominate discussion. They can be asserted when actors use ideas in ‘good currency’ to bolster arguments, or taken for granted when actors operate within ‘paradigms’.

• Shifting policy conditions and events - only some of which are predictable, such as an ageing population or regular elections - can prompt major shifts of policymaker attention at short notice (Cairney and Heikkila, 2014; Cairney, 2015; Hall, 1993; Ostrom, 2007; Weible et al, 2012; Birkland, 1997).

Further, some theories identify complex policymaking systems in which the same inputs of evidence can receive no, or disproportionate attention, and policy outcomes often ‘emerge’ in the absence of central government control (Geyer and Cairney, 2015; Cairney, 2012b; 2015; 2016a: 39; Cartwright and Hardie, 2012: 162-9; Jones and Baumgartner, 2005).

A focus on this bigger picture shifts our attention from the use of evidence by an elite group of elected policymakers at the ‘top’ to a wide range of influential actors in a multi-level process. Many actors present evidence to secure a policymaker audience. Support for evidence-based solutions varies according to which organisation takes the lead and how it understands the problem. Some networks are close-knit and difficult to access because bureaucracies have operating procedures that favour particular sources of evidence and participants. The language in networks takes time to learn. It indicates which ways of thinking about a problem and its solution dominate discussion. Well-established beliefs provide the context for policymaking: new evidence on the effectiveness of a policy solution must be accompanied by a shift of attention and successful persuasion. Sometimes, evidence of socio-economic ‘crises’ prompt lurches of attention. Or, since policymaker attention lurches from issue to issue, there are ‘windows of opportunity’ to present scientific evidence at the right time (Kingdon, 1984: 173).

3. Policymakers combine several principles to produce ‘good policymaking’

Policymakers may value governance principles as much as evidence of problems and solutions (Cairney, 2016b). They refer to several principles when seeking ‘good policymaking’, and only one is EBPM. They may seek policy consensus, to: reflect the general value of pragmatism and cooperation in politics (Lindblom, 1959, 1964, 1979); combine ‘expert scientific advice with a responsiveness to public values’ (Jasanoff, 1986: 5; Weale, 2001: 414); and, improve policy delivery by generating ‘ownership’ of policy among key stakeholders (Richardson and Jordan, 1979; Jordan and Maloney, 1997; Cairney, 2012a: 90). Central governments also share responsibility and accountability with local policymakers, recognising: more than one electoral mandate, the importance of partnerships between local public bodies and stakeholders, and the benefits of tailoring policy to local communities.

The pursuit of several principles affects the balance between national and local policymaking and evidence use (Cairney et al, 2016b). National governments make reference to uniform
delivery standards, to avoid a ‘postcode lottery’, and encourage local autonomy and policy flexibility (Cairney, 2016c). In doing so, note the potential for debates about how we gather and interpret evidence. In some fields, such as health and public health, there is a culture built on EBM principles and a hierarchy of evidence feeding directly into relatively uniform local practice. In others, scholars challenge such principles (Pawson, 2006: 52–4), or combine arguments on evidence and decentralised governance to support the generation of knowledge from local practitioners, service users, interest groups, and public ‘deliberation’ (Williams and Glasby 2010: 97). Consequently, for some scholars there appears to be a large gap between scientific evidence and policy when national policymakers do not adhere to principles of EBM, but not all scholars – and few policymakers – refer to the same standard.

Categories of EBPM and PBE

This discussion helps us separate egregious examples of PBE, driven by ideology and looking for any supporting information, from decisions based on a more sincere combination of evidence, principles, and judgement. As table 1 suggests, the former refers to three of the twelve listed scenarios (in italics), compared with three or four examples of EBM-style EBPM (the shaded area). It leaves us with many cases which are not so easily labelled PBE but might be criticised because policymakers draw on judgement and principles, economic considerations, and/or evidence derived from methods low on an evidential hierarchy.

Table 1: scenarios of EBPM and PBE

<table>
<thead>
<tr>
<th>The scientific evidence on problems and solutions come first, then policymakers:</th>
<th>select the best interventions according to narrow scientific criteria (e.g. on their effectiveness)</th>
<th>base their decision primarily on economic factors such as value for money (VFM)</th>
<th>have ideological, electoral, and/or principled reasons for favouring non-evidence-based interventions.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policymakers recognise a problem demonstrated by scientific evidence, then select a solution built on:</td>
<td>evidence (on its effectiveness) from randomized control trials.</td>
<td>evidence (on its effectiveness) from qualitative data (feedback from service users and professional experience)</td>
<td>their personal experience and assessment of what is politically feasible, and/or democratic or governance principles</td>
</tr>
<tr>
<td>Policymakers identify a problem, then select a solution built on</td>
<td>evidence (on its effectiveness) from randomized control trials</td>
<td>evidence (on its effectiveness) from qualitative data</td>
<td>their personal experience and assessment of what is politically feasible, and/or democratic or governance principles</td>
</tr>
</tbody>
</table>
Policymakers identify a problem: decide what they want to do, then seek evidence to back up their decisions.

the evidence base is not well developed, but policymakers act quickly anyway (and e.g. inflate the likely success of interventions)

the evidence is highly contested, and policymakers select the recommendations of one group of experts and reject those of another.

Note: this list could expand to include combinations, such as when a programme is built partially on promising but incomplete evidence (e.g. from pilots), then rejected or supported completely when events prompt policymakers to act more quickly than expected. Or, an evidence-based programme may be imported without evaluation to check if it works as intended in a new arena. The VFM column is partly shaded because cost-effectiveness is a key part of the calculation in reviews but policymakers often use cruder financial measures.

In other words, one-third of these categories refer to the inevitable role of political judgement: to identify good-enough explanation from the available evidence in a limited time, and weigh it up with experience and governance principles to produce a feasible strategy. This process opens up a wide range of contributors to a policy decision, including sources – such as quantitative and qualitative evidence, expert advice, and/ or counterfactuals – that are often valued more highly by professions and disciplines outside of EBM.

UK Government ‘families policies’: evidence based policy or policy based evidence?

When viewed chronologically from the most recent events, the UK Government’s ‘troubled families’ programme almost seems like the ideal-type PBE in which ministers identified a problem, made a decision based on minimal evidence, and sought to generate evidence to demonstrate success. The programme involves a massive ‘roll out’ of initiatives to intervene in the lives of particular families to (a) address issues such as anti-social behaviour, criminality, child truancy, and parental worklessness, and (b) prevent outcomes such as family eviction, by (c) providing support and threatening sanctions for non-engagement. It piloted several programmes from the late 1990s and promised their expansion under Labour governments before the Conservative-led government announced their massive expansion - to almost 120000 families from 2012-15 and to 400000 from 2015 – as its main solution to the 2011 riots in England (Hayden and Jenkins, 2014).

Within one week of the riots, then Prime Minister David Cameron (2011a) linked behaviour directly to ‘thugs’ and immorality - ‘people showing indifference to right and wrong…people with a twisted moral code…people with a complete absence of self-restraint’ – before identifying a breakdown in family life as a major factor. Cameron (2011b) stressed the need for people to take moral responsibility for their actions, and for the state to intervene earlier in their lives:

Officialdom might call them ‘families with multiple disadvantages’. Some in the press might call them ‘neighbours from hell’. Whatever you call them, we’ve known for years that a relatively small number of families are the source of a large proportion of the problems in society. Drug addiction. Alcohol abuse. Crime. A culture of disruption and irresponsibility that cascades through generations. We’ve always known that these families cost an extraordinary amount of money…but now we’ve come up the actual
figures. Last year the state spent an estimated £9 billion on just 120,000 families…that is around £75,000 per family.

The policy appears to be riddled with PBE (Gregg, 2010; Crossley, 2015a), but three main practices fit into different categories: its identification of the problem and evaluation of its success seems cynical, but different to its interpretation of neuroscientific evidence to frame the case for early intervention, and of the evidence for success of family interventions.

*Its identification of the most troubled families and measure of success in ‘turning them around’*

At the heart of the programme is the assertion that we know who the ‘troubled families’ are, what causes their behaviour, and how to stop it. Yet, much is built on value judgements about feckless parents and tipping the balance from support to sanctions, and anecdotes about the tendency of ‘worklessness’ or ‘welfare dependency’ to pass down generations (Crossley, 2015b; MacMillan, 2014a; 2014b). The government also appears to conflate criteria to identify families in need of *support*, such as the mental health of the mother, and *sanction*, such as criminality or anti-social behaviour (Gregg, 2010: 14; Garrett, 2007).

The government’s target of almost 120000 families was based speculatively on previous Cabinet Office estimates in 2006 that about ‘2% of families in England experience multiple and complex difficulties’ (Kendall et al, 2010: 1; Cabinet Office Social Exclusion Task Force, 2007: 4; National Audit Office, 2013: 5; Hayden and Jenkins, 2014: 635). This estimate was based on limited survey data and modelling to identify families who met five of seven criteria relating to unemployment, poor housing, parental education, the mental health of the mother, the chronic illness or disability of either parent, an income below 60% of the median, and an inability to buy certain items of food or clothing (Levitas, 2012: 4; Hayden and Jenkins, 2014: 635). It then gave estimates to each local authority and asked them to find that number of families, identifying households with:

1. A child who has committed an offense in the last year, or is subject to an anti-social behaviour order (ASBO).
2. A child excluded from school permanently, or suspended on three consecutive terms, in a Pupil Referral Unit, off the school roll, or has over 15% unauthorised absences over three consecutive terms.
3. An adult receiving out of work benefits.

If the household met all three criteria, they would be included automatically (Crossley, 2015b: 3; National Audit Office, 2013: 5). Otherwise, local authorities had the discretion to identify further troubled families meeting two of the criteria and other indicators of concerns about ‘high costs’ of late intervention such as, ‘a child who is on a Child Protection Plan’, ‘Families subject to frequent police call-outs or arrests’, and ‘Families with health problems’ linked to mental health, addiction, chronic conditions, domestic abuse, and teenage pregnancy (DCLG, 2012: 5). Finally, it offered local authorities up to £4000 per family (some up front, some after indicating success) if they invested a further £6000 in effective interventions (while cutting local authority funding in other areas).
Its measure of success, in ‘turning around’ troubled families, was more problematic. It declared almost-complete success without meeting the standard of official statistics (DCLG, 2015). Success ‘in the last 6 months’ is measured by: (1) the child no longer having three exclusions in a row, a reduction in the child offending rate of 33% or ASB rate of 60%, and/or the adult entering a relevant ‘progress to work’ programme; or (2) at least one adult moving from out of work benefits to continuous employment (2012: 9; Casey, 2014: 61; compare with Danil, 2013). So, the success of a policy to address multiple challenges is measured according to change in one (National Audit Office, 2013: 6). It was self-declared by local authorities – albeit subject to DCLG ‘spot checks’ - and both parties had a high incentive to declare it: local authorities received per-family payments and the government received a way to declare progress without long term evidence. The declaration contrasts with an allegedly suppressed report stating that the programme had ‘no discernible effect on unemployment, truancy or criminality’ (Cook, 2016). The latter was partly confirmed by evidence that: many families received no intervention but showed improvement anyway (Bawden, 2016), local authorities could only identify families by departing from the DCLG’s initial criteria (Levitas, 2014; Crossley, 2015b: 6; 2016; Hayden and Jenkins, 2014: 641), only a tiny minority of local authorities invested more than £4000 per family, and there was huge local variation in performance (National Audit Office, 2013: 7-9). There is certainly no evidence to support the government’s heroic claim that spending £10000 per family saves £65000.

Its use of neuroscientific evidence on the benefits of early intervention

The government’s interpretation of neuroscience seems to reflect the existence of some important but highly unclear evidence translated into a politically feasible strategy. It is driven partly by a belief in the benefits of early intervention in the lives of children (from 0-3). Policymakers interpret neuroscientific evidence to emphasise the profound effect of stress and neglect on early brain development. This interpretation contributes to the ‘now or never’ argument found in the Munro review (2011: 69-70). The case is made most vividly by the Allen reviews’ (2011a: 1; 2011b: 1) use of the now-famous images of the brains of ‘normal’ and ‘extremely neglected’ three year old children.

Such neuroscientifc accounts are critiqued heavily in fields such as social science, neuroscientifc, and psychology (Rose and Rose, 2016; Featherstone et al, 2013:5; see also Rutter, 2002). Wastell and White (2012) find no good quality scientifc evidence behind the comparison of child brain development reproduced in Allen’s reports. They suggest that this image is used primarily for its shock value. The UK appeared to follow the US example in which neuroscience ‘was chosen as the scientifc vehicle for the public relations campaign to promote early childhood programs more for rhetorical, than scientifc reasons’ (Bruer, 2011: 2; 1999; see also Shonkoff and Bales, 2011; Gillies, 2014).

Its investment in programmes with limited evidence of success

The government’s use of evidence on previous programme success seems like the weakest candidate for PBE: it identifies a problem and builds policy on the available qualitative evidence. Its expansion of ‘family intervention projects’ (FIPs) to address ‘problem’ or
‘troubled’ families began under Labour. Gregg (2010) describes the expansion as a ‘classic case’ of PBE based on limited evidence of FIP success from a small sample of people from a small number of pilots. Fletcher et al (2012) describe the evidence for FIP effectiveness as ‘weak’, referring to a government-commissioned systematic review which suggests that there are no good quality evaluations to demonstrate the effectiveness or value-for-money of its key processes (Newman, 2007). The impact of related interventions has also been limited, such as the Family Nurse Partnership (FNP) which has so far produced ‘no additional short-term benefit’ (Robling et al, 2015). On that basis, Crossley and Lambert (2016) suggest that “the weight of evidence surrounding ‘family intervention’ and similar approaches, over the longue durée, actually suggests that the approach doesn’t work”.

Yet, we need to examine more realistically the likely availability of supportive evidence for expanded programmes. It is generated in two main ways, from the FNP evaluated primarily by RCTs, to FIPs evaluated primarily with qualitative methods such as interviews with service users and practitioners and with reference to case files (see Nixon et al, 2010: 321-2; chapter 5; Cairney, 2015c), supplemented with stories of individuals (DCLG, 2006: 6).

**Developing an evidence base: 1. Learning from UK experience and FIP pilots**

One key approach to policy learning is to draw on UK experiences by piloting and evaluating projects. This approach is exemplified by the *Dundee Families Project* (DFP), established in 1996, which the UK government supported then sought to ‘roll out’ from 2006, built on independent evaluations which present highly qualified but generally positive accounts of early progress (Nixon et al, 2010: 306; Parr, 2009: 1257; DCLG, 2006; Cabinet Office Social Exclusion Task Force, 2008: 9). Before 2006, this expansion was driven largely by partnerships between individual local authorities and third sector bodies such as NCH Action for Children (which delivered the DFP and subsequent programmes in England), funded largely by central government (DCLG, 2006: 3). The DFP focused on low income, often lone parent, families “who are homeless or at severe risk of homelessness as a result of ‘antisocial behaviour’”. It provided 24/7 support, including after school clubs for children and parenting skills classes, and treatment for addiction or depression in some cases, in dedicated core accommodation with strict rules on access and behaviour, or via ‘dispersed tenancies’ or an outreach model (Dillane et al, 2001: v). Although driven by the perceived success of intensive projects, most FIP projects in England have offered outreach rather than residential services (Nixon et al, 2010: 310; DCLG, 2006: 2-3). After funding 53 ‘Pathfinder’ pilots up to 2008 (Cabinet Office Social Exclusion Task Force, 2008), Labour’s proposal was to reach 50000 families by 2009 (Gregg, 2010: 1).

To some extent, FIPs symbolised a combination of values and evidence to prompt a shift in the ASB agenda from enforcement to a ‘twin track’ approach including greater support and a reduction in the use of ASBOs (Parr, 2009: 1262) – albeit, when compared to their Scottish counterparts, UK ministers described the use of FIPs with stronger reference to ASB and the offer of support combined with the threat of sanctions for non-engagement (Nixon et al, 2010: 306; Parr, 2009: 1259; Casey, 2012: 4).
The evidence of FIP success

The DFP was evaluated qualitatively, primarily using a small number of in-depth interviews from a sample of residents and staff (Dillane et al, 2001: vi), supplemented by self-reporting of success by the project’s management (33, or 59% of cases), and estimated savings based on counterfactuals: for example, what would residential childcare have cost if they did not intervene early? The authors described qualified success built on good management and inter-organisational commitment, and the use of ‘specific intervention types that are tailored to individual families’ needs’ (2001: ix).

Subsequently, Pawson et al (2009: 1; see also Nixon et al, 2010: 307) evaluated the DFP as part of expanded provision in Scotland, including the Aberdeen Families Project and three ‘Breaking the Cycle’ pilot projects using ‘outreach support’ only. They report the comprehensive analysis of case study backgrounds and interventions, supplemented by 78 in-depth interviews of members of 51 families, and discuss success in terms of (for example) 70% of cases closed after completing the agreed programme, ‘reduced complaints of anti-social behaviour’ (‘94% of cases’), and staff assessment of the reduced likelihood of poor outcomes such as homelessness (81%) and continued drug addiction (2009: 5). Pawson et al (2009: 6) use a similar ‘cost consequences’ calculation to highlight some notional short-savings savings, but without the ability to make a definitive judgement.

The review of initial FIP expansion in England - supporting 256 families (370 adults, 743 children) - makes bolder claims about programme success, arguing that ‘the projects had helped them achieve remarkable changes’, including at least 80% of cases in which families were no longer vulnerable to eviction, coupled with ‘significant improvements in children’s health, well-being and educational attainment’ (DCLG, 2006: 7).

In each case, unlike in key programmes evaluated with RCTs testing the effect of a specific intervention, the evaluations of FIP pilots did not provide a ‘blueprint’ or model for emulation (2006: 7). Nor do they describe anything but the potential for these projects to save large amounts of money by reducing demand in acute services (2006: 8). Instead, they identified something closer to a set of principles of good practice which FIPs should be based, including the need for: many agencies to form partnerships; a long term programme of support (1-2 years if families began in supported accommodation, and over 6 months if supported via outreach); and, an ethos of challenging individual and family behaviour ‘based on the professional values of listening, being non-judgemental, promoting well being, and establishing relationships of trust’ (2006: 7). So, when the DCLG (2006: 8-9) describes ‘a robust evidence base to further the development of the Respect Action Plan’ and roll out of FIPs, it refers to general good practice rather than specific interventions conducive to study with RCTs. Or, proponents value the potential for local variation, and the discretion afforded to support workers to deliver services that do not reflect the government’s ASB and problem family rhetoric (Parr, 2009: 1269-70).

The next roll out in England produced 53 new or modified FIPs supporting 690 families in 2007. The authors of its evaluation made similar qualified claims of success based largely on
interviews with FIP staff – in the 90 studied families, who engaged with the programme and completed it successfully during the evaluation period, ‘ASB and criminal activities had declined considerably’ – and expanded on the elements of good practice underpinning successful FIPs: ‘recruitment and retention of high quality staff, small caseloads, having a dedicated key worker who manages a family and works intensively with them, a whole-family approach, staying involved with a family for as long as necessary, scope to use resources creatively, using sanctions with support, and effective multi-agency relationships (White et al, 2008: 2). They also refer to the limitations of their study, pointing to their inability to measure reductions in ASB (their assessment of progress is based primarily on practitioner judgement – 2008: 40), and absence of RCTs to determine the longer term effectiveness of FIPs (2008: 7). This may be particularly important if the initial pilots prove to be unrepresentative because they are administered by relatively highly paid and qualified support workers (Parr, 2009: 1268).

Similarly, Kendall et al’s (2010: iii) evaluation of ‘Family Pathfinder’ pilots in England use interviews with, and online surveys of, practitioners in local initiatives to identify the potential cost savings of intervention, with an illustrative programme for 53 families suggesting that ‘One million pounds of family intervention costs is estimated to generate savings of £2.5m by avoiding adverse outcomes for family members; a net benefit saving of £1.5m’. In this case, the difference comes from a specific estimate of costs, using the SROI approach (Social Return on Investment - described by Nicholls et al (2009). The authors stress the illustrative and preliminary nature of such estimates, and note that the estimate savings ‘cannot necessarily be cashed by local authorities’ (2010: iii). Rather, ‘the benefit cost savings need to be viewed at a society, rather than a local authority level’. This approach developed some traction in the Treasury, which adapted its PBR (Payment By Results) system to families policy to encourage local authorities to invest and give them extra funding based on estimates of savings to other public bodies (interview, HM Treasury, 2015).

Overall, such evaluations provide highly qualified indicators of success or promise. They compare with key non-commissioned reports, such as by Gregg (2010), who criticises the implication that FIPs turned around the lives of over 80% of the ‘worst’ families when the figures: (a) often related to a small sample of the population, and (b) described people relatively willing to engage with FIPs because they were at risk of eviction. Further, their risk of eviction was often linked to issues of mental illness, debt, and unemployment - ‘hardly the image of ASB fed to the public’ - not solved during the evaluation period (Gregg, 2010: 3-5; 15).

Developing an evidence base: 2. Learning from international experience and RCTs

Another approach is to learn from, and/ or seek to transfer, the success of specific interventions from other countries (Rose, 1993; 2005; Dolowitz and Marsh, 1996; 2000; Berry and Berry, 2007; Cairney, 2012: 244). The UK Government supports a collection of interventions whose success has been generated primarily with reference to EBM, evidence of success from multiple RCTs – initially outside the UK – of interventions requiring ‘fidelity’ to make sure that the ‘dosage’ and its effect can be measured (Oliver et al, 2014; Cairney, 2016a; 2016b).
The Family Nurse Partnership (FNP)

The FNP began in the US as the Nurse-Family Partnership, designed to engage nurses with first time mothers (deemed to be at relatively high risk of poor life chances) approximately once per month from pregnancy until the child is two. The criteria for inclusion relate to age (the UK focus is mainly teenage pregnancies), income (low), and partnership status (generally unmarried). Nurses give advice on how mothers can look after their own health, care for their child, minimise the chances of further unplanned pregnancy, and access education or employment. The FNP combines an intervention to address the immediate problems faced by mothers and ‘early intervention’ to influence the longer term impact on children. It is based on several theories of human behaviour:

- Attachment theory is the basis for nurses aiming to develop a therapeutic alliance with mothers, with many activities introduced to enhance understanding of how to develop mother-infant and father-infant relationships; self-efficacy is introduced to help women to gain control over their lives in their relationships and in life course planning; and ecological theory underpins the timing of the intervention at the point of an ecological transition, and in the attention paid to enhancing family support and links with community services (Barnes, 2010: 9).

This programme gained its reputation from RCTs which often demonstrated high effectiveness and low cost. The US-based Coalition for Evidence-Based Policy (2012) gave it ‘top tier’ status, which describes ‘Interventions shown in well-designed and implemented randomized controlled trials, preferably conducted in typical community settings, to produce sizable, sustained benefits to participants and/or society’. It describes common outcomes in at least two US RCTs, including reductions in pre-natal smoking, child abuse and neglect, and second pregnancies, and improvements in their child’s cognitive function and education attainment in follow-up studies (when the children reached 15-19) at a low cost.

It is in this context that the UK Government imported the FNP: establishing an internationally recognised evidence base using EBM criteria and with reference to bodies such as the UK’s Cochrane Collaboration or US’ Society for Prevention Research; and, seeking to replicate the programme in England (Barnes, 2010: 5).

The Department of Health gauged its ability to replicate the US programme with a pilot, initially of 10 sites, followed by semi-structured interviews with 77 FNP staff and, for example, structured interviews with an 8-10% sample of service users (Barnes, 2010: 16). The pilot identified potential problems with implementation, such as when deciding how to recruit a cohort of the most vulnerable service users, with some sites focusing exclusively on age (below 20) and others extending the age to 24 and adding criteria: ‘not in education, employment or training (NEET), or no educational qualifications, or no supportive partner’ (2010: 16). However, during policy delivery, most respondents appeared to be positive about engaging with the programme, ‘not perceiving that they had been identified as parents likely to fail but as parents who would benefit from much needed support’ (Barnes, 2010: 16).
After piloting, the programme was rolled out in England to 9000 expectant mothers, with reference to its high cost effectiveness and ‘strong evidence base’, which would be enhanced by an RCT to evaluate its effect in a new country (Family Nurse Partnership National Unit, 2014; Barnes, 2010: 10). Crucially, the FNP requires fidelity to the US programme (you can only access the programme if you agree to the licensing conditions) based on evaluation results which showed that the programme was most effective when provided by nurses/ midwives and using a license ‘setting out core model elements covering clinical delivery, staff competencies and organisational standards to ensure it is delivered well’ (Department of Health, 2012: 6). Fidelity is a requirement because, ‘If evidence-based programmes are diluted or compromised when implemented, research shows that they are unlikely to replicate the benefits’ (2012: 6; Barnes, 2010: 9), and the FNP website outlines ‘fidelity goals’ which resemble those for prescription medicines. In practice, this produces a ‘fair degree of fidelity’ (Barnes, 2010: 27). Indeed, although Olds (2015) was involved to some extent in the English process, he argues that some problems with initial selection of service users is contributing to the disappointing results of the first RCT (Robling et al, 2015).

**Triple P (Positive Parenting Program)**

Triple P began in Australia, as a parenting programme with five levels to reflect severity of need, from ‘community information provision to intensive one-to one work’ (Lindsay et al, 2011: 3). The Triple P website describes it as a ‘parenting and family support system designed to prevent – as well as treat – behavioural and emotional problems in children and teenagers’ (Triple P, 2016). It offers a standard programme for children up to 12, plus specialist courses for parents of, for example, 12-16 year old, disabled, and overweight children. It emphasises high flexibility based on levels of intensity of intervention, evidence-based effectiveness, and low cost, to describe Triple P as an intervention that can be delivered as whole population or targeted programmes (by trained practitioners using the same manual).

Its claim to be one of the most evidence-based interventions in the world has proved contentious. Wilson et al’s (2012) review of 33 studies finds limited effectiveness at often high cost, and there has been some debate between the authors of studies promoting Triple P versus those who question its cost-effectiveness (Tellegen and Sofronoff, 2015; Reijneveld et al, 2015; Coyne, 2015; Sanders et al, 2012; Coyne and Kwakkenbos, 2013). In Scotland, evaluators did not recommend the continuation of Triple P without some form of RCT to demonstrate its value (Marryat et al, 2014: 8).

There are similar concerns in England about Triple P effectiveness, partly because evaluations are patchy and provide highly qualified results. Burney and Geldsthorne (2008: 478) described it as a ‘programme much favoured by the government’ before being ‘independently evaluated’. Lindsay et al’s (2011: 3; 12) subsequent evaluation identifies ‘positive changes in the small to medium range for child problem behaviour, parent well-being and parenting skills’ but also the absence of an RCT and some difficulties in gathering data. Therefore, non-imported programmes ‘might be equally effective’ (Churchill and Clarke, 2010: 49). Lewis (2011: 107) notes that courses are generally provided or commissioned by local authorities (as opposed to primary care settings in Australia) and that they are expensive to provide. The Early
Intervention Foundation (2016: 106-9) provides heavily qualified support for specific versions, while noting uncertainty about the cost.

**Incredible Years**

Incredible Years began in the US (developed by Professor Carolyn Webster-Stratton) as a training programme ‘for families with severely behaviourally disordered children’ (0-8 years). It uses a written curriculum, media, and short workshops to ‘teach parents how to manage difficult behavior’ (Waldfogel and Washbrook, 2011: 7), train teachers to develop effective classroom techniques, and/or treat ‘clinic-referred’ children or whole classrooms (Bywater and Sharples, 2012: 397). Overall, it has a ‘strong evidence base’ (2011: 7; Lindsay et al, 2011: 3). Incredible Years appears to be the only intervention of the three to have received favourable evidence from UK RCTs (three in England and Wales), and is supported by NICE as part of the Improving Access to Psychological Therapies programme in NHS England (EIF, 2016: 103-4).

Notably, the specific application Incredible Years Preschool BASIC (target-indicated) has the highest (level 4+) rating from the Early Intervention Foundation (EIF, 2016: 97), ‘meaning that it has evidence from over three RCTs demonstrating short-term improvements in children’s behaviour’ (one study highlights benefits over ten years). It also receives a low-medium cost rating (2). Designed for ages 3-6, it involves one two-hour session per week for approximately 20 weeks, in which two trained practitioners use media and group work to encourage parents to ‘learn strategies for interacting and communicating positively with their child, promoting optimal social and emotional development and discouraging unwanted child behaviour’ (the next stage, ADVANCED, addresses issues such as ‘anger and depression management’ - 2016: 103). There is also some scope for additional phone, email, or home visit follow ups (combined with at-home exercises) in individual cases.

Most notably, the programme with the strongest evidence of short and long-term success is target-indicated, which means that it was administered to parents of children already deemed in need of treatment. There is limited evidence of success in ‘target-selective’ interventions based on identifying high risk from factors such as socio-economic conditions, partly because ‘parents may not have felt so motivated to continue to work on their child’s behavior, since it was less problematic at the outset and since they were not seeking treatment when recruited’ (Scott et al, 2014: 655; see also EIF, 2016: 105-6; Statham and Smith, 2010: 5; these problems seem to be magnified in whole population initiatives – Boffley, 2016). This experience reinforces a wider dilemma in early intervention: there are few examples of taking effective projects ‘to scale’; there are major issues around ‘fidelity’ to the original project when you scale up (including the need to oversee a major expansion in well-trained practitioners); and, it is difficult to predict the effect of a programme, which showed promise when applied to one population, to a new and different population (Dodge, 2009).

**Discussion: two stories from the same evidence**

If we reverse the chronology of these decisions – to focus on the initially available evidence – and consider the political context in which Westminster governments have to make and defend
policy, we can identify a different, central government, perspective. Policymakers have to make choices in the face of uncertainty: they recognise the limits to existing data, but need to choose quickly, especially if there is no realistic prospect of ever receiving completely supportive evidence. In Westminster systems in particular, they perceive the need to make choices unequivocally and demonstrate success to protect themselves and their investment in policy. Central governments have to project an image of control because they know that other actors will try to hold them to account in elections and debate. This dynamic has an enduring effect on the design and delivery of public services regardless of the commitment of governments to encourage delegated governance. While prevention and early intervention policy is vague, individual programmes such as ‘troubled families’ contain enough detail to generate intense debate on government policy and performance. So, they contain elements which emphasise ‘muscular, effective government’ (Davies, 2015: 17), including sustained ministerial commitment and a determination to demonstrate early success.

This context helps us compare three stories, from most unsympathetic to most sympathetic to the government’s motives. The former suggest that the government is responsible for egregious PBE, inventing statistics to declare the success of a programme built on arbitrary and stigmatising measures of a problem, and providing no neuroscientific or evaluation evidence to justify such a massive expansion of a failed programme. It often dominates academic discussions, combining empirical evaluation with normative criticisms of the stigmatising effects of policy (Parr, 2006: 1260; Garrett, 2007; Gregg, 2010).

The latter can be derived from a central government perspective: the government found a way to turn limited but broadly supportive evidence into a high profile commitment to preventive spending and early intervention. Early intervention represents a heuristic: policymakers think that it is important to their values, they generally receive good feedback on this approach, there is some evidence to support a causal link between childhood neglect or trauma on poor life chances and, for example, the unequivocal or exaggerated interpretation of vivid neuroscience may represent the best way to sell policy change. These ends (good preventive policy) justify the means (defending policy with limited or problematic evidence), using the rhetorical value of neuroscience and political crises to encourage rapid policy change. This national agenda, combining government frameworks and funding incentives, provides ‘cover’ for local action. Local authorities retain the discretion to commission the most ‘evidence-based’ interventions, and practitioners can modify FIPs in light of professional values, under the cover of a high priority programme in which artificial short term evaluation aids long term progress.

The main effect of the policy is to invite local authorities and their partners to fund or deliver two types of family intervention or parenting programme designed to intervene as early as possible in people’s lives to improve their life chances. It is possible for this process – of setting national direction but encouraging local use of well-regarded projects - to become more ‘evidence based’ than it first appears, since public bodies may choose programmes with reference to promising evaluations.

Viewed in this way, local governments have a suite of options which demonstrate their evidence base or effectiveness in different ways. The FIPs are built on a combination of
principles of ‘good practice’ to solve individual problems such as risk to housing tenure: having a specific worker assigned to specific families, providing ‘hands on’ support to intervene directly in their choices, challenging their views about their problems, considering the effect of each family member on the other, and coordinating multi-agency partnership working (Department for Communities and Local Government, 2012: 6-8). They are evaluated using case reports and surveys and interviews with support workers and parents, largely with the aid of counterfactuals (this programme is expensive but crucial, and the consequences of non-intervention would have been more expensive).

Programmes such as the FNP, Triple P, and Incredible Years are justified and evaluated in different ways – generally from international RCTs followed in some cases by domestic RCTs – but with similar levels of uncertainty about their effectiveness in local areas, or in wider populations, across the UK. In fact the EIF (2016: 11) states that it is, ‘inappropriate to draw strong conclusions about which programmes will work or will not work when each programme only has a small number of evaluations and few have very rigorous or long-term evaluation across multiple sites’. Further, the intervention with the most impressive evidence (Incredible Years) seems only to be effective during a tertiary stage of prevention, rather than in primary/universal interventions or a secondary process of identifying high risk.

Therefore, in each case, the identification of evidence does not act as a substitute for choice. Central government appears content to set a national framework and give some choice to local public bodies, backed increasingly by advice from the (mostly government funded) Early Intervention Foundation which maintains a database of evidence-based programmes and a star-rating system measuring effectiveness and cost. In practice, this approach limits central direction, and there appears to be no central record of local choices (EIF, 2016: 20).

Consequently, there is also a third story which combines elements of the other two: scholars raise major concerns about the nature and tone of government policy before describing a tendency for policy to change as is implemented, such as when mediated by local authority choices and social workers maintaining a commitment to their professional values when delivering policy (Featherstone et al, 2013: 7; Morris and Featherstone, 2010; Hayden and Jenkins, 2013: 468; Danil, 2013; Butler, 2014: 420).

**Conclusion**

PBE may be a dramatic political slogan, but it does not provide a useful empirical description of the many choices that combine to produce ‘policy’. As an umbrella term, it conflates a range of practices, from a cynical process to decide first and justify later, to the routine need to make choices based on principles, values, and judgement despite high levels of scientific uncertainty. The alternative is to use policy studies to interpret the process partly from the perspective of policymakers facing: their own bounded rationality, major limits to the influence in complex policymaking systems, and the need to justify policies only partly with reference to evidence.

This perspective allows us to categorise more effectively the use of evidence in policy, and consider the motives of policymakers. So, in ‘troubled families’ policy, we can see three main examples: the problematic development of indicators of ‘troubled families’ and policy success;
the exaggeration or misleading use of neuroscientific evidence to justify early intervention; and, the use of highly qualified evidence of promise to justify the massive expansion of family intervention and other projects. From a critical outsider’s perspective, this order is important: ministers decided what they wanted to do without supportive evidence to back up their policy, so developed ridiculously misleading, and potentially stigmatising, measures of the problem and success of the solution. From an elected policymaker’s perspective, the order and interpretation of events is different: they began with evidence that was promising under the circumstances but never likely to provide a ‘magic bullet’, made a judgement on expansion (and some commitment to measure progress via bodies such as the EIF), and used the tools of government necessary to sell and defend policy in high stakes Westminster systems.

In such debates, it is difficult to separate ideological from empirical evaluations. Two actors can agree completely on the evidence base – on the size of the problem and effectiveness of existing solutions - but disagree completely on its implications for future policy, which involves deciding how we should describe and treat people, how much money we should spend on particular programmes, and the value for money of those programmes. In such cases, focusing solely on the extent to which policy is ‘evidence-based’, or identifying PBE too broadly, can downplay the importance of the politics of evidence-based policymaking in all policymaking. No policy can, or should, be based entirely on evidence, and we need categories and standards to reflect this point.

References


Burney, E. and Geldsthorne (2008) “Do We Need a ‘Naughty Step’? Rethinking the Parenting Order After Ten Years”, The Howard Journal, 47, 5, 470–485


Cairney, P. (2015a) “What is ‘complex government’ and what can we do about it?”, *Public Money and Management*, January, 3-6


Family Nurse Partnership National Unit, 2014, The evidence base for Family Nurse Partnership


MacMillan, L. (2014a) “While there is evidence that workless spells are associated across generations, the ‘culture of welfare dependency’ argument does not hold” http://blogs.lse.ac.uk/politicsandpolicy/intergenerational-worklessness-the-unexplored-curse/


http://bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-10-130