

**Integrating quasi-experimental and inductive designs in
evaluation: a case study of the impact of free bus travel on
public health**

Judith Green*, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, UK

Helen Roberts, UCL Institute of Child Health, UK

Mark Petticrew, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, UK

Rebecca Steinbach, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, UK

Anna Goodman, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, UK

Alasdair Jones, Department of Methodology, London School of Economics, London, UK

Phil Edwards, Faculty of Epidemiology and Population Health London School of Hygiene and Tropical Medicine, UK

***Corresponding author Address:** Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London WC1H 9SH **Tel: 0044 (0)20 7927 202**
Email: Judith.green@lshtm.ac.uk

Note: this is a personal version, created by Anna Goodman, of the text of the accepted journal article. It reflects all changes made in the peer review process, but does not incorporate any minor modifications made at the proof stage. The 'in press' citation for the final journal article is:

- *Green, J., Roberts, H., Petticrew, M., Steinbach, R., Goodman, A., Jones, A. & Edwards, P. In press. Integrating quasi-experimental and inductive designs in evaluation: a case study of the impact of free bus travel on public health. Evaluation. 21: 391-406*

<http://evi.sagepub.com/content/21/4/391.full.pdf+html>

Copyright © and Moral Rights for this paper are retained by the individual authors and/or other copyright owners

Abstract

Evaluations of ‘natural experiments’ in public policy are typically considered ‘weak’ evidence. Challenges include: making credible claims for causal inference (internal validity); generalising beyond the case (external validity); and providing useful evidence for decision makers. In public health, where experimental evidence is encouraged by funders and enjoys a degree of rhetorical favour, in theory if not practice, current guidance for evaluating natural experiments focuses largely on methods for strengthening *internal* validity. Using a case study of the evaluation of free bus travel for young people in London, UK, we demonstrate a pragmatic approach to strengthening both internal and external validity in evaluations through integrating the logic of quasi-experimental methods with inductive qualitative analysis. Combining theoretical and inductive analysis in this way to address questions of policy interest through evaluations of natural experiments may be fruitful, and have methodological advantages over randomised designs.

Résumé

Les évaluations des ‘expériences naturelles’ dans le domaine des politiques publiques sont souvent considérées comme des preuves peu fiables. Les principaux défis incluent: Émettre des affirmations crédibles afin de dériver inférences causales (validité interne); Généraliser au-delà du cas observé (validité externe); Et fournir des preuves utiles aux décideurs. Dans le domaine de la santé publique, où les preuves expérimentales sont encouragées par les investisseurs et bénéficient d’une certaine faveur (du moins en théorie) les principales orientations relatives à l’évaluation des expériences naturelles se concentrent principalement sur des méthodes visant à renforcer la validité interne. En nous basant sur une étude de cas concernant l’évaluation de la gratuité des transports en bus pour les jeunes gens à Londres, nous proposons une approche pragmatique renforçant la validité interne et externe des évaluations à travers l’intégration d’une logique quasi-expérimentale et celle d’une analyse qualitative inductive. Associant une approche théorique à celle d’une analyse inductive afin de répondre aux problèmes que posent l’évaluation des politiques publiques peut être productif et peut aussi présenter des avantages méthodologiques par rapport aux études randomisées.

Key Words: natural experiment, analytic induction, public transport, policy evaluation, public health.

Policy evaluation and natural experiments

A growing commitment in many policy arenas to ‘evidence informed’ public policy (Nutley, et al 2008) has been accompanied by considerable debate about the relative merits of experimental and more ‘realist’ approaches to evaluation (Byrne, 2013; Marchal et al., 2013). Randomised controlled trials (RCTs) are frequently inappropriate or impossible for evaluating complex interventions in domains such as public health, transport or housing (Petticrew et al., 2005; Petticrew and Roberts, 2003), and are typically used in few evaluations (Woolcock, 2013). However, in public health in particular, RCTs remain, normatively, the gold standard design for evaluating effectiveness. In this context, natural experimental studies have attracted renewed methodological attention largely in terms of attempts to defend them against the threats to rigour arising from the non-random allocation of interventions. Current guidance on methods to enhance causal credibility in evaluations of natural experiments therefore focus largely on threats to internal validity, and on potential study design modifications and statistical analyses to address issues such as selection bias and confounding (Craig et al., 2012). From this perspective, the challenge of quasi-experimental design is that of replicating, as far as possible, the strengths of randomised designs, in which causation can be established through probabilistic comparisons of how discrete variables are associated with outcomes.

However, threats to internal validity are not the only methodological challenge of quasi-experimental designs. Often based on single case studies, evaluations of natural experiments are also subject to the same potential weaknesses of limited generalisability that apply to ‘true’ experimental designs in which intervention allocation has been randomised. The most rigorously designed trial can only demonstrate that ‘the intervention’ had effects within the trial population at a particular point in time. No matter how carefully the sample is drawn, causal claims cannot easily be extrapolated from the study population to any larger one without understanding the processes which give rise to the relationships between ‘intervention’, ‘population’ and ‘outcome’. Feedback loops are likely to shape the causal pathways within the system (Shiell et al., 2008), and the system is likely to have emergent properties which cannot be predicted by aggregating individual-level effects (Byrne, 2013). All these factors pose challenges when it comes to answering the crucial question for policy evaluations aiming to inform practice; namely, ‘will it work here?’ (Arai et al., 2005; Cartwright, 2007; Cartwright and Munro, 2010). To assess whether an intervention implemented in one setting is likely to have similar effects elsewhere, decision makers need evidence on the underlying principles by which an intervention works and on the key contextual factors which influence its effects. For these reasons, there have been trenchant critiques of the prioritisation of experimental design in general in evidence based policy, with calls for developments of case-based analyses such as Qualitative Comparative Analysis to strengthen transferability (Byrne, 2013; Blackman, 2013; Blackman et al., 2013; Woolcock, 2013).

A less well-documented pragmatic challenge for evaluation of natural experiments in public policy is identifying the outcomes of most importance to the various constituencies

with an interest. For a transport policy intervention with health impacts, for instance, these constituencies might include: national and regional transport and public health commissioners and providers; businesses affected by transport; individuals and communities affected by the policy; and tax payers. Researchers primarily interested in the health outcomes of transport policies have to accommodate the reality that these are rarely the primary or only outcomes of interest to those designing, commissioning, or implementing transport interventions. For a public health audience, the most credible evaluation designs focus on the health outcomes of interventions that can be clearly delineated from the messy policy environment in which they were implemented. Designs in which ‘confounders’ can be excised, and ‘fidelity’ assured, are (in principle) those in which the causal relationship between intervention and outcome can be most credibly inferred. However, the context in which ‘real world’ policy interventions happen is important for understanding effects, both in terms of the ways in which it shapes the meanings of an intervention for given populations, and how it constructs implementation in practice. In methodological terms, tight control therefore potentially trades credibility with policy decision makers (‘this looks like the kind of thing that might be useful for our setting’) for high internal validity (‘we can be sure that the effect measured resulted from our intervention’). For fields such as public health, such trade-offs (Mercer et al. 2007) are a significant challenge for developing a useful as well as robust evidence base, simply because the ‘tightest’ evaluation designs are often the easiest to implement for the most ‘downstream’ of interventions. The further upstream in policy terms the intervention is, the more difficult it is to delineate single interventions and outcomes. In a review of transport and public health, for instance, Ogilvie and colleagues (2005) found that the ‘strongest’ evidence (most likely to come from randomised designs) came from largely ‘downstream’ behavioural interventions, whereas larger scale policy interventions (such as city wide environmental regeneration programmes) had only ‘weaker’ evidence from descriptive case studies. What they call the ‘inverse evidence law’ is worrying in a context of evidence-informed policy making, given that it is these more ‘upstream’ interventions which are most likely to have an impact on the determinants of health and health equity (Lorenc et al., 2012).

One approach to both including the range of outcomes likely to be of interest to different constituencies and avoiding trade-offs between external and internal validity is to mix designs and data collection methods in an iterative way. Ton (2012), for instance, describes a pragmatic approach drawing on realist evaluation, which begins with a focus on the questions (and outcomes) most relevant to the client, and then anticipates a range of potential threats to validity (and thus credibility). This paper builds on these pragmatic approaches. We outline how integrating qualitative analytic induction within a quasi-experimental design can help defend against threats to validity whilst ensuring that the evaluation results have purchase with a wide range of constituencies. In doing this, we draw on Nancy Cartwright’s suggestion that, in evaluating programmes, as well as asking does ‘X cause Y?’, it might also be productive to ask ‘how ‘stable’ is the capacity of X in effecting a change in Y?’ (Cartwright and Munro, 2010). A focus on the ‘stability’ of the capacities of policies and programmes does not preclude attention to questions of causation and impact, but it does involve asking a somewhat broader question about ‘what happens’ when a policy is introduced. This requires marshalling a broad range of

evidence to develop a credible argument about the likely limits and possibilities of transferability.

Given that the criteria for assessing the stability of capacities are not as well developed as the criteria for assessing the causal claims in RCTs, there is a requirement to demonstrate not only that X causes a change in Y but also to examine *how* this comes about (the ‘mechanisms’ and pathways in epidemiological terms). This entails setting out what other factors can promote, inhibit or otherwise affect the ability of X to effect a change in Y (Cartwright and Munro, 2010). Thus, suggest Cartwright and Munro, an evaluation might ask a set of questions such as: How does X operate to promote Y?; What is needed for X to promote Y?; What can stop/inhibit the operation of X?; What other capacities promote/inhibit Y?; What happens when capacities interact? Here, we take a case study of an evaluation of a natural experiment to demonstrate how drawing on a range of study designs conceptualised within a logic model (Baxter et al., 2010) can shed light on these questions of the capacity of an intervention to cause changes in the determinants of health, and do so without trading internal and external validity.

The case study: free bus travel for young people in London, UK

Our case study is an evaluation of a policy ‘natural experiment’, the introduction of free bus travel for young people in London, UK, in 2005/6. In many settings with public transport systems, young people can travel using concessionary fares. A number of schemes also provide subsidised fares to particular population groups to enhance social inclusion, such as for young people with disabilities or those from lower income groups (Grayling 2001). The scheme introduced in London in 2005 was unusual for a concessionary fare scheme in that it provided free (to the user) travel on all greater London’s bus network for all young people of school age, regardless of financial or other need. This was extended to all of those under 18 in full-time education or training in 2006. London’s transport infrastructure and policy environment is unusual. Apart from Northern Ireland, London is the only region within the UK with a regulated bus system, with Transport for London (TfL) holding responsibility for implementing the (elected) Mayor of London’s transport strategy. Transport for London retain responsibility for planning bus routes, managing service quality and providing infrastructure such as bus stops and shelters, contracting with (largely) private operating companies to run bus services.

Offering free travel for young people was part of a raft of policy initiatives, with aims including the encouragement of public transport and discouragement of car use in the city. Other initiatives included: expansion of the bus network; a congestion charge zone, implemented in 2003, which has reduced car journeys within the central area (Prud’homme and Bocarejo, 2005); and various policies to encourage cycling. Transport for London’s own stated aims for the free bus scheme were “to help young people to continue studying, improve employment prospects and promote the use of public transport” (Transport for London, 2006:7). These environmental and social inclusion aims can in themselves be considered ‘upstream’ interventions for public health, with potential future implications for young people’s access to determinants of health such as

education and training, and for their contribution to reducing pollution if free bus travel does reduce private car use. However, these were unlikely to have been the only health effects. Public debates around the implementation and future of the scheme suggested a number of other anticipated implications, both positive and negative, for the health and wellbeing of young people, and for others affected by their travel. In the context of rising childhood obesity levels and policy efforts to get children walking and cycling more, politicians and public health practitioners raised concerns about the potential impact on young people's levels of 'active travel' if they had no disincentive to use the bus for even the shortest of journeys (Evans, 2011; GLA, 2010). Other concerns were raised in policy briefings and the media, including: the possible impact of young passengers limiting the access of older bus users, either through physical displacement or 'fear based exclusion' (London Assembly Transport Committee, 2008); the impact on crime on transport (Transport for London 2008); the vulnerability of young people to assault; and the cost-effectiveness of a scheme open to all young people regardless of financial need.

Assessing the potential risks and benefits for different elements of health and wellbeing was difficult in the absence of evaluations from similar schemes, particularly given that few urban administrations face similar challenges to those of London. In settings in which private car use is the most viable alternative to bus travel, improvements in public transport may increase the amount of 'active travel', simply because reaching a transport hub will involve some walking (Besser and Dannenberg, 2005; Coronini-Cronberg et al., 2012). However, in many European cities, where walking or cycling are realistic alternatives to bus transport, reducing the financial barriers to public transport may reduce the amount of active travel undertaken. Some 'evidence' was therefore needed to inform debate about the scheme, but to be useful, any evaluation would ideally shed light on the general issue of the circumstances in which concessionary fare interventions are likely to affect determinants of public health, as well as identifying specific effects of the London scheme.

The problem: a 'messy' intervention in a complex system

The free bus scheme, was then, potentially complex, given that it was likely to interact with a raft of other policies designed to change transport mode choices in London, national transport policy interventions and broader social change. Given London's unique transport infrastructure (particularly the investment in bus transport), there were no obvious comparison populations of young people. With no control, potential confounders in any before and after design range from the theoretically knowable (e.g. other interventions also potentially influencing the behavioural outputs of interest, such as the introduction of congestion charging) and the more general and difficult to measure cultural shifts, such as rising concern about children's independent mobility, or obesity. Many potential confounders remain unknowable in such observational studies. The causal pathways that link the intervention (providing free bus travel) and the outcomes of health policy interest (health behaviours and outcomes) are long and likely to behave in ways that characterise any complex system. That is, their component parts may interact (synergistically or otherwise), and the pathways may be non-linear, and subject to feedback loops, phase changes and tipping points. As an example, providing free travel

might, over time, devalue the status of bus travel or result in uncomfortably overcrowded buses, thus reducing uptake after an initial rise in use. Finally, different evidence users are interested in different outcomes: road safety engineers, for instance, may prioritise injury outcomes over longer term health benefits from active travel.

A solution: investigating the ‘capacity’ of the intervention

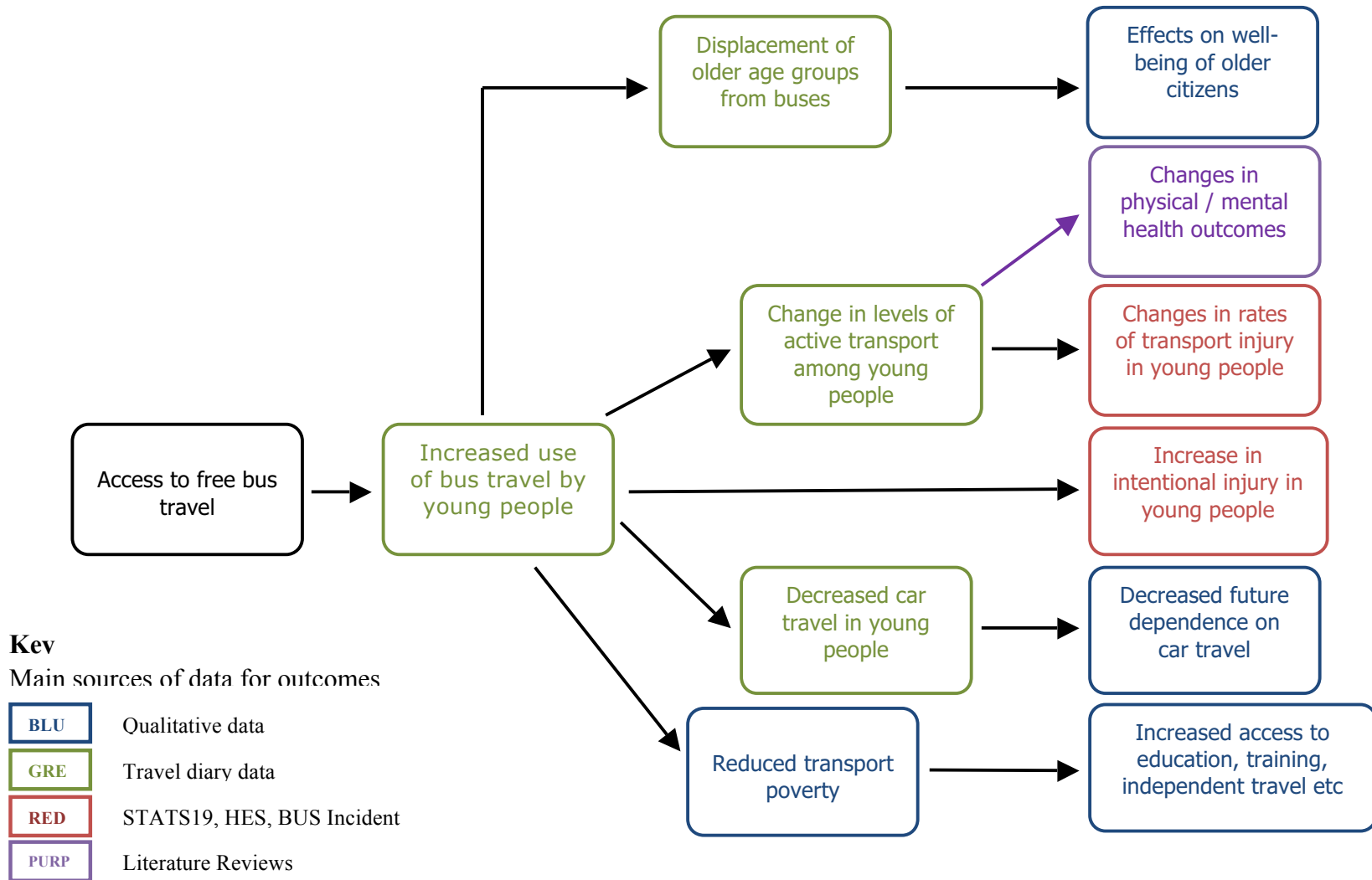
To address these threats to internal validity, external validity, and credibility for evidence users, we designed an evaluation of free bus travel which drew on three major elements: a logic model, which mapped out the main pathways of interest to policy makers and practitioners; an epidemiological study utilising current best practice in the evaluation of natural experiments (Craig et al., 2012) to identify changes over time along these pathways that could plausibly be attributed to the intervention; and a substantial qualitative component (using both deductive and inductive methods) designed to: contribute to the epidemiological study (that is, contribute to a variable-based study); to understand why the impacts identified in quantitative data sets did or did not happen (understanding mechanisms); and to understand the part played by (free) bus travel in young people’s wellbeing (that is, contribute to the study of the intervention as ‘a case’). Combining these methodological approaches, we argue, enabled us to provide ‘good enough’ evidence on the impacts of the scheme which were credible for policy makers. Crucial to the process was the development of a logic model at the design stage that addressed the broader question of the ‘capacity’ of free bus travel to affect health outcomes, and that was revised at the end of the project. A logic model is a graphic depiction of the pathways that are hypothesised to link the policy intervention with the outcomes of interest. This paper summarises how the findings of the epidemiological and qualitative elements of our evaluation were brought together in the light of our initial logic model in order to provide as robust an evaluation as possible of the health impacts of the scheme, and one designed to be useful to decision makers in London and beyond. Detailed methods and findings of elements of the study are published elsewhere (Jones et al., 2012; Edwards et al., 2013; Goodman et al., 2014; Green et al., 2014a, 2014b). Here, we draw on selected findings to demonstrate how the qualitative analysis contributed to addressing some of the methodological and knowledge transfer problems generated by policy evaluations in general, and by evaluations of natural experiments in particular. We focus particularly on examples to show the role of inductive qualitative analysis in the elucidation of the capacities of the intervention. Illustrative extracts from the qualitative data are from individual or group interviews indicate gender (M or F) and age or age range (for group interviews) of the speaker, but have otherwise been anonymised to protect confidentiality.

Designing the evaluation: mapping outcomes, indicators, hypothesised causal pathways and data sources

The first step was to map the hypothesised relationships between free bus travel and the health outcomes of interest in the light of theory, the academic literature on transport interventions and the views of evidence users, including transport commissioners and providers, public health practitioners and transport users. Key variables on this causal

pathway were then specified in order to identify sources of data. The main ones were graphically depicted on a summary logic model (see Figure 1). The development of the logic model places theory at the heart of the evaluation as it makes explicit the theoretical and other assumptions underlying the evaluation. The purpose of the evaluation then becomes the building of a plausible model of what happens in selected parts of a complex system over a particular time period. We hypothesised that offering free bus travel would: increase access to transport for young people; increase the overall number of trips made; increase social inclusion; increase road traffic injury (as a result of increasing exposure to road danger); and increase vulnerability to assault.

Figure 1: Initial Logic Model



A precondition for epidemiological evaluations of natural experiments are that there are data available on ‘exposed’ and ‘non-exposed’ populations before and after the intervention (Craig et al., 2012). For many of our outcomes of interest existing data sets included sufficient ‘before’ and ‘after’ data to identify significant changes in behaviour or outcome before (non-exposed) and after (exposed) the scheme was introduced (Edwards et al., 2013). These included: travel-diary surveys conducted by Transport for London to measure travel mode usage; police records of road injury; and hospital records of admissions for assault injuries. To strengthen our ability to make causal inferences about the effect of the intervention, the main analyses used a change-on-change analysis to provide some control for background changes in travel behaviour over the period of the study by comparing changes in target population (young people) with changes in adults aged 25-59 (i.e. those old enough never to have experienced free bus travel as a young person, and too young to qualify for an older person’s free bus pass) (Edwards et al., 2013).

Some outcomes of interest were rather less easy to operationalise and measure. The explicit aims of the scheme were to reduce future dependence on car travel and to address transport poverty for young people. Although current and past car occupancy can be captured by travel diary data, future dependence cannot. ‘Transport poverty’ is a rather abstract concept (Church et al., 2000) which can be operationalised in a number of ways, including access and ability to pay for transport. No sufficiently detailed quantitative datasets were available which directly measured these outcomes, and instead a qualitative study was designed to generate indicative data. However, these qualitative data (derived from individual and group interviews and observations of the transport system) also provided a rich source of evidence for analytic induction, enabling us to treat the ‘intervention’ as a case. This enabled us to begin to delineate the necessary and sufficient conditions for the effects identified in the epidemiological change-on-change analysis. We now outline the ways in which this helped provide evidence on the ‘capacities’ of the intervention, focusing on outcomes related to active travel and transport poverty.

Qualitative analysis to help understand how ‘X operates to promote Y’

An important outcome for public health is the impact of free bus travel on ‘active travel’, a concept originally operationalised in our logic model as ‘distance walked and cycled’. Analysis of travel diary data suggested that, relative to adults unaffected by the scheme, young people were: travelling further by bus, making an increased number of short trips by bus, and reducing the number of kilometres travelled by cycle. However, there was no significant decrease in the number of kilometres travelled by walking (Edwards et al., 2013).

The findings that many more short trips (potentially walking trips) are now made by bus and that there was a significant decrease in young people’s cycling (compared with adults) could suggest a causal link between the intervention and decreased ‘active transport’. A thematic analysis of the qualitative data initially appeared to support this, given that examples of using the bus as a default mode of transport for even the shortest of journeys permeated the data. To take just one illustration, this young man’s description

of his journey to school is extreme, but suggests the lengths young people apparently go to avoid walking:

My dad takes me a couple of metres down the road, it's only about 200 metres down the road. And then from then I go and get the bus to school. And then there's only a few metres from where I get off the bus to go to school. ...I'm on the bus for roughly about less than a minute. (M, 14)

Further, had the qualitative data been used simply for triangulation, young people's explicit rationales for their frequent short trips by bus would have suggested that these do indeed result from the fact that they have free travel:

I think that the fact that knowing the bus is free helps me want to get on the bus... if I had to pay I would not get on the bus. (F, 16)

However, a more inductive approach to the qualitative data, using it to understand the *capacity* of free bus travel to affect active transport, begins to identify a rather different perspective. First, as suggested by quantitative data, bus travel of course generates a considerable number of 'trips', all of which entail a certain amount of activity:

If I didn't have free travel...I wouldn't be going places I would be probably staying quite local and through using free travel it means I can go places that I've always wanted to go (M, 15)

This is in line with mobility research which suggests that a considerable amount of population activity is achieved in the context of transport itself (Besser and Dannenberg, 2005). Second, ethnographic data from participant observations of travel suggest that bus trips are not necessarily sedentary for young people (Jones et al., 2012). Indeed, the journey home from school in particular, is a fairly 'active' event, with physical activity on the bus, between buses and around bus stops. Operationalising 'active travel' as simply the number (or proportion) of trips or kilometres travelled by walking and cycling may be misleading.

On cycling, although the quantitative data suggests a decline associated with the intervention, it is clear from analysis of young people's accounts that cycling is simply not considered a viable transport mode in the capital: it is framed primarily as a leisure activity, associated with either 'hanging out' or with childhood:

Interviewer: Do you ever ride a bicycle?

...I love riding bikes ... for fun ... I think I'm too old for it now (M, 18)

I used to cycle every Sunday with my dad and younger brother, but now we don't anymore. Yeah, just, we kind of stopped, yeah. (F, 15)

Although cycling rates did fall in young people relative to adults post-intervention, it is difficult to argue that this is a direct result of free bus travel, simply because cycling is rarely framed as an alternative candidate mode of transport for this age group. This is in part because a key driver of travel mode choice for young people is sociability: travelling together was almost always the primary aim, implicit or explicit of transport choice in our

discussions. This is to offset boredom, to minimise risks and to maximise the opportunities for socialising with peers:

[L]ike we'll just be bored and we don't want to go home, so we'll just hop on a bus and we'll go anywhere. (F, 17).

Cycling, in a context where few young people routinely use cycles for transport, is not conducive to sociability in the way that walking or taking the bus can be. In this context, free bus travel replaces walking for many short journeys because it can also be used to 'travel together', since everyone is entitled to a free bus pass (Goodman et al., 2014). Thus, an important element of 'the intervention' emerged. This was the fact that it is universal which enabled everyone in a peer group to travel together: this is a necessary condition of the bus becoming the default mode of transport.

An explicit aim of the scheme was to discourage future car dependence (Transport for London, 2006). This is impossible to evaluate directly in the short term, and whilst survey data can be used to gauge intentions, these tend to be a rather poor guide to future behaviour. Again, a thematic analysis of the content of the qualitative data suggested that young people were still firmly committed to future driving, as a rite of passage, a route to independence, freedom and adult identity. In response to direct questions about future intentions to learn to drive, typical exchanges focused on the 'adult' rewards that might accrue from driving:

M: As soon as I'm old enough to drive I'm going to drive!

M: ... when I come home from parties I'm cold... so cold, so I always imagine I have a car just come...

M: Yeah, a car you could drive!

M... also it's more impressive to girls if you've got a car, if you're 20 and you haven't got a car and you

M just do public transport

M: can't go out to meet girls and that ...

M: If you're still rocking an Oyster [public transport pass]!

All: [laughter] (M, 14-15)

However, taking such utterances as 'indicators' of future car use does not provide much in the way of useful information: they are perhaps more validly read as indicators of the ways in which driving is still associated with adult status and independence. A more inductive analysis of data reveals a number of implicit assumptions about bus travel made by young people. A key one is that it is the 'taken for granted' way to get around, evidenced first in the ways that young people easily described other passengers using a range of typologies suggesting almost every population group: workers, old people, business people, mothers. Bus travel was the default mode for young people, often explicitly preferred to lifts from parents because it required no parental funding or permission, and could include all of a peer group (Green et al., 2014a). In London at least, this suggests, 'everyone' travels by public transport. Since it does not carry the same stigma in London as in some other major cities, it may be that habituation to using the bus may decrease the likelihood of future car use.

The qualitative analysis therefore has two functions in helping to understand how free bus travel affects ‘active travel’ outcomes. First, it can unpack the mechanisms: for instance in suggesting how bus travel apparently replaces walking, but does not (overall) reduce the amount of walking by shedding light on the ways in which young people both travel more overall with a bus pass and the ways in which they envisage other forms of travel such as cycling or being driven. Second, it helps refine the links between concepts and the indicators: here, how ‘active travel’ is operationalised as ‘trips by walking and cycling’. These both require an analysis which goes beyond simply reporting respondents’ views.

Strengthening credibility of claims about ‘what is needed for X to promote Y’

Our original hypothesised causal pathways included a link between more bus travel, reduced transport poverty and increased access to education or training. These are important pathways in that they reflect the aims of the scheme, and a core determinant of health and health inequalities is education. The quantitative data suggested little evidence for this causal pathway, with no evidence for a declining social class gradient in the number of trips to school and college over time (Edwards et al., 2013). However, these data were restricted to analysis of school term time weekday travel, which may have underestimated the effects of the scheme on transport poverty.

Qualitative data provided some evidence for the scheme having an effect on transport poverty and education at the margins, for those living in families where free bus travel made an important direct contribution to household budgets. However, two further findings may be more significant. First was the complete absence of comments in the data set about financial barriers to transport: young people from all backgrounds discussed travel as a taken-for-granted given. In general, they could travel where and when they pleased:

I take the bus every day... [for] going to school, going to dancing, going to see my friends, maybe going to church... because it’s free ... I can go to different places, so anywhere I want to go (M, 15)

[The scheme is] really helpful, whenever I really need to go anywhere it’s just, it’s no hassle (M, 14-18).

We had no comparative populations to make the case that ‘transport poverty’ has disappeared in London *because* of the free bus scheme – it may well be that discounted child bus fares in other cities have the same effect. However, the exceptions (deviant cases) in our data set strengthened the case for the free bus scheme having some causal effect. These were those who had had the card providing free travel confiscated:

[W]hen I didn’t have [free bus travel] I did struggle in terms of not getting everything done because I didn’t have that freedom to get on a bus (M, 12-17).

[of friends who’ve had card confiscated] It puts a strain on their social activities because they can’t go out as much (M, 15).

That the free scheme (rather than any fare reduction) was crucial (i.e. causal) to eliminating transport poverty was evident in the exceptions. Those young people who had lost their card or had it confiscated, for instance for infringing the behaviour code, were the only able-bodied young people to report transport related restrictions on their mobility.

The case that this absence of transport poverty has benefits for health, even if it has no apparent influence on gradients of access to education, is strengthened by analysis of the role that travel plays in young people's lives. Journeys to school or college might be prioritised in household expenditure and thus paid for by parents, however constrained their circumstances, and therefore there may be little observable impact from the scheme on the social gradient of journeys to school and college. However, discretionary journeys, which parents may be reluctant to pay for, emerge as vital for young people's social inclusion. They are first a rare space for young people to socialise:

It's one of the main things you do on the bus, if you go out with someone you sit down and you talk about things. (M, 14-18)

Second, in a city with restricted opportunities for young people to independently engage in activities outside the home which are not organised by adults, 'catching the bus' is one of the few activities that is affordable to an entire peer group:

Mostly every Saturday [my friend and I] will probably just jump on a bus, because we have a free [travel], and go anywhere and get another bus from there, and another one. And we just travel, we don't know where we're going, we just jump on a bus because we can. (M, 15)

So if we ask 'what is needed for X to promote Y?', or what is needed for a free bus scheme to influence health through the pathway of removing transport poverty, one key necessary condition would appear to be a *universal* free bus scheme (Goodman et al., 2014). Sociability is key to young people's travel decisions, and bus use, if free for all, becomes the 'lowest common denominator', because it is available to all a peer group:

My mum or dad would drive me if I want them to but it's like I said you meet friends on the bus and things like that. (M, 13-16).

These advantages only work if everyone can travel together.

What inhibits X promoting Y?: the lack of an accessible bus service

A second exception in the data helps to unravel what inhibits free bus travel from having positive effects on health through the pathway of inclusion. This was one group of young people, those with disabilities, whose accounts of using buses were completely different from the rest of the sample. For them, buses were not a facilitator of transport, sociability and inclusion but rather a barrier to getting things done. Although London buses have ramps for wheelchairs, we were told that these are often not working; that you might have to wait a long time for a bus with sufficient space for a wheelchair; that buses might be accessible, but this was not always the case for bus stops; and, importantly, that there is room for only one wheelchair at a time on most buses, which precludes social travel with other wheelchair users. Thus, although young people with disabilities also have access to

free travel, buses were inconvenient and anti-social and no disabled young people reported using buses as a site for socialisation, or for a 'fun' day out.

Some ramps don't tend to work, so that's a bit of a hassle ... sometimes it's dangerous with an electric chair, it's heavy... (M, >16)

The intervention (free bus travel) therefore only has the capacity to act (to improve the ability to travel together, to improve transport access) if there is an available and accessible bus network.

Discussion

This paper has discussed some effects identified from an evaluation of the natural experiment of free bus travel for young people in London, UK. We have summarised evidence (see Green et al., 2014a for more detail) that it enhanced social inclusion without reducing the amount of active travel; and that it made a contribution to 'destigmatising' bus transport, an important precondition of reducing private car use. We argue that these effects are likely to hold in other settings where there is an efficient and accessible bus service (i.e. one that is perceived as offering a density of routes and frequent reliable services) and where the scheme is a universal, rather than conditional, entitlement. To make these inferences, we drew on a mix of data sources to elucidate the *capacities* of an intervention (free bus travel for young people) to affect public health. A quasi-experimental design using change-on-change analyses (comparing changes in exposed and non-exposed populations before and after the intervention) provided evidence of effects that were associated with the intervention. Qualitative evidence was used to enhance causal credibility and judgements about external validity.

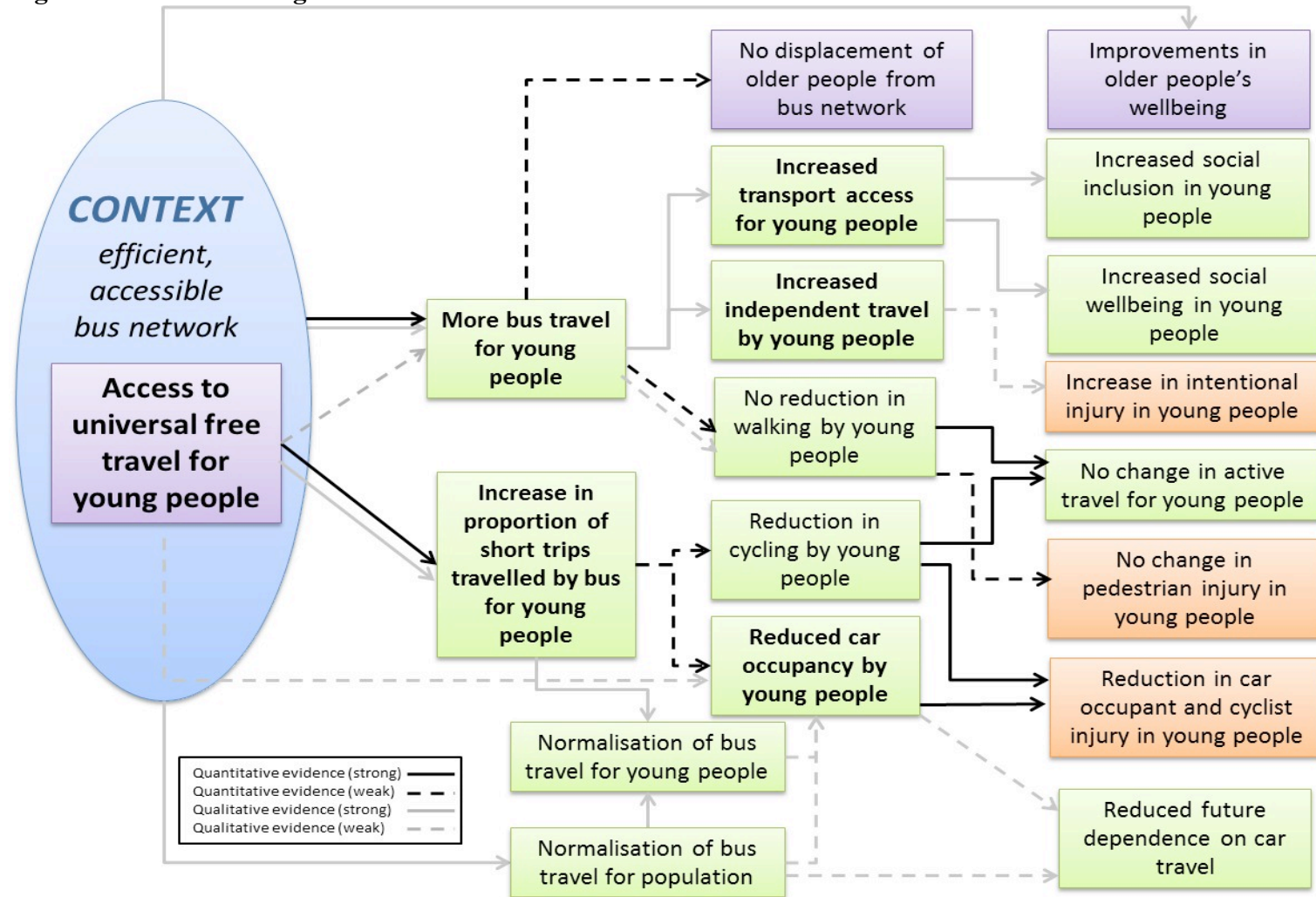
The qualitative data has been used to 'make sense' of otherwise counterintuitive findings (e.g. that more bus travel did not necessarily mean less active travel), and to exercise caution in attributing causal effects to other associations (e.g. that cycling reductions are a causal result of the intervention). Analyses of qualitative data provided two distinct contributions. First, they are 'weak', or suggestive, evidence, for some of the health effects that are hypothesised, but which cannot be directly measured, such as future intentions to drive and thus the impact on the sustainability of the transport system. More importantly, however, the second contribution, from an inductive analysis, was to strengthen the credibility of causal claims within the case, and to strengthen the claims we could make about the transferability of impacts identified.

What might have happened had we simply used the qualitative data for triangulation, or for unpacking the processes of the intervention? We have suggested that starting with the epidemiological data, and trawling the qualitative transcripts for data relating to key 'themes', or for what young people had to *say* on the topics emerging from the quantitative data would have led to misleading findings, such as that the scheme itself had reduced cycling rates, or that it had had no impact on propensity to drive. Instead, a qualitative analysis using more inductive logic which starts with the question 'what is going on here?' and analyses the data in its own terms, as well as in terms of categories arising from the epidemiological model, provides greater explanatory purchase. This

suggests that qualitative research has a key role in understanding capacities in practice. While qualitative research retains an important place in relation to understanding the salience and acceptability of interventions and barriers and levers to their success, it can also be pivotal in enhancing the credibility of research evidence on complex interventions. This includes roles for qualitative research in enhancing understanding of the operation of the intervention in context, and in relating theoretical concepts of interest to the empirical indicators utilised in (in particular) secondary data analysis.

Our findings were integrated into a revised logic model (see Figure 2), which incorporated both the evidence from quantitative methods (which used secondary data sets to compare changes before and after the intervention in the target population to those in other populations) and the inductive qualitative work.

Figure 2: The revised logic model



This revised logic model (see Figure 2) can now specify the intervention, in that it can delineate what is necessary for it to have the capacity to effect the changes we identified in our case. That is, for free bus travel to reduce transport poverty, have no impact on active travel and to contribute to reduced future car dependence requires that the scheme is universal, and that it is implemented in the context of a good and accessible bus service. This specification is what aids those in other jurisdictions in evaluating whether these conditions are likely to hold. Further comparisons with other cases (such as conditional concessions in other cities) will enable these claims about transferability to be extended and strengthened, through the use of, for instance, Qualitative Comparative Analysis (Ragin, 1999; Blackman, 2013; Blackman et al., 2013). However, we have shown how comparisons *within* the case can also be used to aid the causal claims made.

Evaluations of natural experiments have been considered a poor relation to RCTs for evaluating policy; utilised when there is no choice, and to be defended against threats to internal validity through approximating the advantages of randomisation as far as possible. However, following Cartwright and Munro (Cartwright 2007; Cartwright 2013; Cartwright and Munro 2010) we have shown that if we add a question relating to the capacity of the intervention to our evaluation, we can strengthen the credibility of causal inference without jeopardising questions of transferability. Indeed, we would suggest that there are situations in which evaluations of natural experiments may be a stronger design than those reliant on randomising cases. In this particular case, had free bus travel been implemented in a trial design with, for instance, random allocation of free bus passes to some low income children, or to particular geographical clusters, it may not have been possible to identify the advantages that accrue specifically from this being a universal benefit, available to all of a peer group. Although (in principle) randomising large units (such as cities) may have captured this, the point is that the emergent property of this system (which relates to the universality of the benefit to a group, not the material benefit per se) would not necessarily have been known prior to randomisation to inform the choice of scale. Studying a natural experiment allows the researchers to identify policies as ‘changes to systems’ with multiple outcomes across that system, rather than ‘changes in systems’, with the burden of trying to identify discrete and independent outcomes.

There will never be a ‘strong’ evaluation of every complex intervention, but most complex interventions have common challenges. These include: providing credible evidence for a range of possible constituencies; defending against threats to internal validity and providing accounts of the effects of the intervention which are likely to be useful beyond the local case. We have suggested that a ‘capacity’ approach to evaluations of natural experiments in public policy can meet these challenges within one case. To some extent, this approach integrates ‘variable based’ and ‘case based’ (Ragin 1999) evaluative strategies. In summary, the approach entailed:

- 1) Drawing on research evidence to develop a logic model of the main pathways linking the intervention to outcomes of interest to stakeholders;
- 2) Identifying the best available evidence within a quasi-experimental design for demonstrating causal relationships on these pathways where these were short enough to be credible;

- 3) Using inductive qualitative analysis to:
 - a. Strengthen the credibility of causal inferences drawn from quantitative data;
 - b. Develop our understanding of causal chains at higher level of abstraction;
 - c. Addressing essential questions about the pathways on the logic model, including: how X promotes Y; what facilitates this relationship; what inhibits this relationship.
- 4) Integrating (2) and (3) within a revised logic model which:
 - a. Refines the characterisation of ‘the intervention’: that is, identifies the necessary and sufficient characteristics for it to have had the identified effects on health outcomes
 - b. Make explicit the range of outcomes that accrue from the policy intervention.

Cartwright (2013) argues that assessments of external validity and predictions about future effectiveness rely on knowing the supporting factors that allow an intervention to have effects both within the intervention setting, and elsewhere. This case study shows how this has been done using a range of qualitative and quantitative data in the context of a logic model.

Conclusion

To inform public policy, and enhance generalisable learning from evaluations in single settings, we need to maximise both internal and external validity. Including qualitative research, and logic models to bring to the surface underlying theory and assumptions, are not just useful ‘added extras’ but are integral to evaluating natural experiments. In fact, it is because of a natural experiment’s observational nature – its ‘naturalness’ – that the credibility of any causal inferences depend crucially on being rooted in a theory of change and an understanding of the pathways derived from appropriate evaluation data. Using a case study of free bus travel for young people, we have demonstrated that natural experiments drawing on secondary analysis of existing data sets of ‘exposed’ and ‘non-exposed’ populations and inductively analysed qualitative evidence in the context of theory-informed logic models, can offer both credible and useful evidence on the impact of public policies on health, and the capacities of those policies in specific contexts.

Conflict of interest

None

Acknowledgements

This paper draws on a project funded by the National Institute for Health Research Public Health Research programme (project number 09/3001/13). The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the Department of Health. AG contributed to this work while funded by a post-doctoral fellowship funded by the NIHR. We would like to thank the anonymous reviewers for helpful comments on an earlier version of this paper.

References

- Arai L, Roen K, Roberts H and Popay J (2005). It might work in Oklahoma, but will it work in Okehampton?. *Injury Prevention*, 11: 148-151
- Baxter S, Killoran A, Kelly MP and Goyder E (2010) Synthesizing diverse evidence: the use of primary qualitative data analysis methods and logic models in public health reviews. *Public Health*, 124: 99-106
- Besser LM and Dannenberg AL (2005) Walking to public transit: steps to help meet physical activity recommendations. *American Journal of Preventive Medicine*, 29: 273-80.
- Blackman T, Wistow J and Byrne D. (2013) Using qualitative comparative analysis to understand complex policy problems. *Evaluation*, 19: 126-140
- Blackman T. (2013) Rethinking policy-related research: charting a path using qualitative comparative analysis and complexity theory. *Contemporary Social Sciences* 8: 333-345
- Byrne D (2013) Evaluating complex interventions in a complex world. *Evaluation*, 19: 217-228
- Cartwright N (2007) Are RCTs the gold standard? *BioSocieties*, 2: 11-20.
- Cartwright N (2013) Knowing what we are talking about: why evidence doesn't always travel. *Evidence and Policy*, 9: 97-112.
- Cartwright N and Munro E (2010) The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice*, 16: 260-66.
- Coronini-Cronberg S, Millett C, Lavery AA and Webb E (2012) The impact of a free Older Persons' Bus Pass on active travel and regular walking in England. *American Journal of Public Health*, 102: 2141-2148
- Craig P, Cooper C, Gunnell D, Haw S, Lawson, K, Macintyre S, Ogilvie D, Petticrew M, Reeves B, Sutton M and Thompson S (2012) Using natural experiments to evaluate population health interventions: new MRC guidance. *Journal of Epidemiology and Community Health*, 66:1182-86
- Church A, Frost M and Sullivan K (2000) Transport and social exclusion in London. *Transport Policy*, 7 (3): 195-205.
- Edwards P, Steinbach R, Green J, Petticrew M, Goodman A, Jones A, Roberts H, Kelly C, Nellthorp J and Wilkinson P (2013) Health impacts of free bus travel for young people: evaluation of a natural experiment in London. *Journal of Epidemiology and Community Health*, 67: 641-647
- Evans R (2011) Scrap free bus travel for kids, Boris – it's fueling obesity. *Local Transport Today*, 566.
- GLA (2010) *Report of Transport Committee Seminar: 'The Future of London's Buses'* London: London Assembly Transport Committee/ Greater London Authority.

- Goodman A, Jones A, Roberts H, Steinbach R and Green J (2014) "We can all just get on a bus and go": rethinking independent mobility in the context of the universal provision of free bus travel to young Londoners. *Mobilities*, 9:275-293
- Grayling T (2001) *An introduction to bus policy in Britain*. In Grayling T (ed) *Any more fares? Delivering better bus services*, London: IPPR.
- Green J, Steinbach R, Jones A, Edwards P, Kelly C, Nellthorp J, Goodman A, Roberts H, Petticrew M, and Wilkinson P. (2014) On the buses: a mixed method evaluation of the impact of free bus travel for young people on the public health. *Public Health Research*, 2: 1
- Green J, Jones A and Roberts H. (2014) More than A to B: the role of free bus travel for the mobility and wellbeing of older citizens in London. *Ageing and Society*, 34:472-94
- Jones A, Steinbach R, Roberts H, Goodman A and Green J. (2012) Rethinking passive transport: bus fare exemptions and young people's wellbeing. *Health and Place*, 18: 605-612
- London Assembly Transport Committee (2008) *Crime and disorder on London's buses*. London: GLA
- Lorenc T, Petticrew M, Welch V and Tugwell P (2012) What types of interventions generate inequalities? Evidence from systematic reviews. *Journal of Epidemiology and Community Health*. 67: 190-193
- Marchal B, Westhorp G, Wong G, Van Bell S, Greenhalgh T, Kegels G, Pawson R (2013) Realist RCTs of complex interventions – an oxymoron. *Social Science and Medicine*, 94: 124-128
- Mercer S, DeVinney B, Fine L, Green L and Dougherty D (2007) Study designs for effectiveness and translation research: identifying trade-offs. *American Journal of Preventative Medicine* 33: 139-154
- Nutley S, N Walter and Davies HTO [eds] (2008) *Using evidence: How research can inform public services*. Bristol: Policy Press
- Ogilvie D, Egan M, Hamilton V and Petticrew M (2005) Systematic reviews of health effects of social interventions: 2. Best available evidence: how low should you go? *Journal of Epidemiology and Community Health*, 59: 886-92.
- Petticrew M and Roberts H. (2003) Evidence, hierarchies and typologies: horses for courses. *Journal of Epidemiology and Community Health*, 57: 527-9
- Petticrew M, Cummins S, Ferrell C, Findlay A, Higgins C, Hoy C, Kearns A and Sparks L. (2005) Natural experiments: an underused tool for public health? *Public Health* 119: 751-7.
- Prud'homme R. and JP Bocarejo (2005) The London congestion charge: a tentative economic appraisal. *Transport Policy*, 12: 279-287.
- Ragin C. (1999) The distinctiveness of case-orientated research *Health Services Research* 34 (5): 1137-1151
- Shiell A, Hawe P and Gold L (2008) Complex interventions or complex systems? Implications for health economic evaluation. *British Medical Journal* 336: 1281-3.
- Ton G (2012) The mixing of methods: a three step process for improving rigour in impact evaluations. *Evaluation*, 18: 5-25
- Transport for London (2006) *Transport for London News*. London: TfL.

- Transport for London (2008) *A problem-oriented policing approach to tackling youth crime and anti-social behavior on London's buses*. London: Transport for London.
- Woolcock M (2013) Using case studies to explore the external validity of 'complex' development interventions. *Evaluation*, 19: 229-248