What caused the drop in teenage pregnancy from 1999 to 2015? A review of reviews of hypothesised causes

# Introduction

As demonstrated in the literature in the previous chapter, observations of high rates of pregnancy amongst teenagers have prompted action to reduce these in developed countries. In the last few decades, large drops in rates have occurred in several of these countries (ref: figure produced in last chapter?). Observers have suggested that several interventions intended to reduce risk of pregnancy amongst teenagers may have each contributed to this overall decrease (for example: Elliott, Henderson, Nixon, & Wight, 2013; Hadley, 2014; Kearney & Levine, 2015; Sipsma, Canavan, Gilliam, & Bradley, 2017; Wellings et al., 2016). Other potential candidates for explaining the effects seen across several countries have also been suggested as global changes to the teenagers’ culture and environment (Girma & Paton, 2015; Kearney & Levine, 2015; Sipsma et al., 2017). Figure 1, produced from the literature, presents some initial hypothesised pathways.

Several examples of policies and interventions, shown in the top left of Figure 1, are known to have been piloted and evaluated, with varying estimates of effect size. Associations with environmental changes across time have also been taken to suggest some of the non-designed causes in the bottom left (for example, Girma & Paton, 2015; Twenge, 2017). However, there may not be sufficient evidence for many causes in both categories to assess the contribution of the hypothesised cause to the observed decreases in rates. Randomised controlled trials can be used to estimate the effects of an intervention on a sample population, but results may not be generalisable to a whole population, or implementation of a previously tested intervention may not achieve the full effectiveness as in the experimental condition [MRC Guidance] (Black, 1996). Additionally, many of the exposures listed in the bottom left of Figure 1 are not able to be applied to or withheld from a population in an experiment, and so cannot be tested in this way. Population-level observations of changes in pregnancy rates associated with possible causes may not be able to adjust for all possible confounders and adequately test causation. However, observations can sometimes produce some confidence in attributing an observed effect to a cause (Glasziou, Chalmers, Rawlins, & McCulloch, 2007).

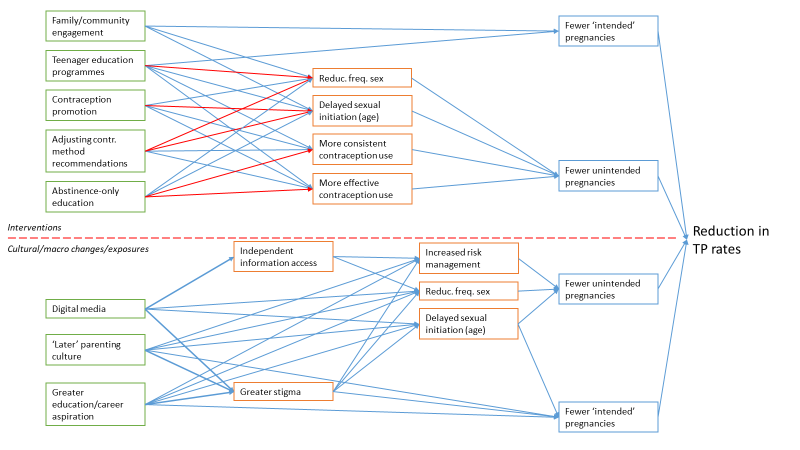


Figure 1 - Theory of change for teenage pregnancy, showing a range of potential causes

The aim of this review is twofold. Firstly, by searching available literature I aim to scope the wide range of suggested causes of reduction in teenage pregnancy presented in the literature with supporting evidence. These may be trials of interventions in experimental situations, evaluations of interventions across larger populations, or observations of changing environments associated with decreasing rates. I restricted searches to published reviews of relevant studies to address the expectedly large volume of literature of all types on the topic. Reviews were taken as likely the most influential publication type on policy decisions and I anticipated that the most important hypothesised causes would be addressed in such reviews. Evidence of the effectiveness of each is assessed to evaluate the plausibility of effects of each. This data is used to update the logic model, removing pathways where the evidence indicates no effect and adding pathways which were not previously known.

Secondly, from this broad list of potential causes, I aim to select those with some degree of plausibility of contribution to the UK’s large declines in rates. I use several methods to assess the strength of evidence, the exposure of the population to the intervention or cultural change and thus the confidence of causal inference using the RE-AIM framework (Glasgow, Vogt, & Boles, 1999).

# Methods

## Aim and Research Questions

The aim of this review was to identify plausible hypothesised causes of decreasing rates of teenage pregnancy, observed in the UK from 1999 to 2015. This was done by answering two research questions:

RQ1: What changes in interventions, culture, policy and environment are hypothesised to cause reductions in teenage pregnancy in the UK and similar countries?

RQ2: Which causes are more likely to have applied to the UK from 1999 to 2015?

## Search strategy

I searched four databases (Medline, Embase, Scopus and Cochrane Database of Systematic Reviews) for reviews of adolescent/teenage pregnancy and prevention or reduction in rates. Where reviews of reviews were found in search results, I manually added all reviewed citations to screening. Full search strategies are presented in [Appendix x].

## Screening

### Inclusion criteria

##### Population

The focus of this review is on female adolescents, aged 13-19 years living in High-Income Countries (HICs) and so experiencing similar political and economic contexts and potentially similar interventions. The selection of countries was taken from two sources. Populations of the 31 countries included in Euro-Peristat reporting (Euro-Peristat, 2018) representing European countries were eligible, alongside four English-speaking HICs: Australia, Canada, New Zealand and the United States of America (see Box 1).

Austria

Australia

Belgium

Bulgaria

Canada

Croatia

Cyprus

Czech Rep.

Denmark

Estonia

Finland

France

Germany

Greece

Hungary

Iceland

Ireland

Italy

Latvia

Lithuania

Luxembourg

Malta

Netherlands

New Zealand

Norway

Poland

Portugal

Romania

Slovakia

Slovenia

Spain

Sweden

Switzerland

UK

United States of America

Box 1 A list of the 35 countries to be included in this review

(compiled from Euro-Peristat countries plus other high-income countries to compare with the UK)

##### Intervention/Exposure

Reviews evaluating interventions which aimed either explicitly to reduce teenage pregnancies or to address associated sexual health risk behaviours were included. Reviews which focussed on the effects of broader cultural-macro changes on pregnancies and risks were also included. These include reviews evaluating interventions which are designed to be applied to a population or a vulnerable subset of a population and are generalisable across national borders.

##### Control

Only reviews reporting a comparison between exposed and unexposed populations were included. Interventions measured either with a control group for medium to long-term outcomes (such as pregnancy) or reporting pre/post measurements of short-term outcomes (such as contraceptive use) were included. Other broader cultural changes measured between populations or across time were included.

##### Outcome

Our primary outcome was rates of pregnancies or births to women aged under 20. Other behaviours directly affecting pregnancy risk – sexual activity and contraceptive use – were included.

##### Study type

Only reviews of other published literature were included.

### Exclusion criteria

I excluded reviews which only looked at abortions, maternal and infant health, and other outcomes after conception. Reviews which evaluated associations of hypothesised determinants of pregnancy risk across a population, with no analysis of changes of exposure over time, were not included. Additionally, reviews which only measured changed attitudes or knowledge as an intervention outcome, rather than behaviour change, were not included, as these were deemed a poor measure of actual risk of pregnancy.

## Analysis 1 – strength of evidence

### Data extraction and synthesis

I constructed a table with a row for each hypothesised cause identified in Figure 1 and extracted from each review a summary of the evidence presented for causes addressed. New rows were added as additional suggested causal pathways were identified.

I coded each review’s evidence for the causes they examined as strong evidence of a positive (or negative) effect on a population’s pregnancy rates, weak evidence of a positive (or negative) effect, or no evidence of an effect.

Evidence was appraised using an adapted set of questions, presented by Howick et al. (2009) as adapted from the Bradford Hill criteria (Hill, 1965). These are arranged to summarise three ‘types’ of evidence, with revised guidelines matching Hill’s originals (Table 1).

|  |  |  |
| --- | --- | --- |
| Type of evidence | Revised, structured guidelines | Hill's original guidelines |
| *Direct* | Size of effect not attributable to plausible confounding | Experiment |
| Appropriate temporal and/or spatial proximity (cause precedes effect and effect occurs after a plausible interval; cause occurs at the same site as the intervention) | Strength |
| Dose‐responsiveness and reversibility | Temporality |
| *Mechanistic* | Evidence for a mechanism of action (biological, chemical, mechanical) | Biological gradient  Biological plausibility |
| Coherence | Coherence |
| *Parallel* | Replicability | Consistency |
| Similarity | Analogy |

Table 1 Bradford Hill's original guidelines and proposed revisions (Howick et al., 2009; Table 1).

'Coherence' is moved from the 'Parallel' subheading to 'Mechanistic', in accordance with the paper's summary of the guidelines

From these, I developed a series of questions to evaluate the evidence given in the reviews for each hypothesised cause. These were separated into two parts, firstly querying the strength of the evidence of effects of the intervention or exposure of interest, and secondly whether the intervention/exposure was shown to be applied to a whole population (or a vulnerable subsection of a population). Whilst the original criteria aimed for specificity, rejecting spurious hypothesised causes, the adapted use of the criteria here aimed for sensitivity, anticipating that strong evidence may not exist for certain causal factors. Additionally, studies may present strong evidence of the effectiveness of an intervention, but the intervention may not have been scaled up to a whole-population application and so could not have contributed to observed reductions.

Reviewed causes were marked with a ‘+’ denoting weak evidence of causing a reduction in pregnancy rates if at least one question regarding evidence strength was answered positively:

1. Is data presented which shows an effect while controlling for plausible confounding? (‘Direct’ evidence – see Table 1)
2. Is data presented which shows an effect whilst NOT controlling, but:
   1. A logic model or narrative synthesis of mechanisms is presented? (‘Mechanistic’); AND
   2. The model is coherent with what is currently known? (‘Mechanistic – Coherence’)
3. Are results seen consistently in different national contexts? (‘Replicability’)

A hypothesis which included at least one of the above, distinguishing it from an observed association with no causal assessment, was increased from ‘weak’ to ‘strong’ evidence if it was plausibly presented as contributing to a whole-population reduction in rates. The following question was used to determine this:

1. Is this hypothesis:
   1. Tested across the population of a whole nation (as included in this analysis – see Box 1); *OR*
   2. Tested amongst a large proportion of the population, or a high-risk group (e.g. targeting prevention of repeat pregnancies or socio-economic subgroup); *OR*
   3. Noted by authors to have been applied to a population other than the observed groups but within this review’s observation period (1990-)? (Direct evidence)

Similarly, if reviews presented evidence of negative effects of an intervention or exposure, increasing pregnancy rates, the evidence was similarly categorised.

### Analysis

After extracting and coding all causes from reviews, I reanalysed these in connection with the previously produced logic model, detecting similarities in interventions or causes which suggested grouping together in branches of the model. I then updated the model adding new branches from newly found suggested causes, adding mediators where appropriate.

Finally, I looked at the evidence presented for each pathway and considered the possibility of its contribution to observed rates. Pathways which represented interventions with some evidence of effectiveness were kept in the model, whether tested across a whole population or not. Similarly, pathways which were untested or had not been addressed in the literature, particularly environmental causes which were expected to be less tested, were kept for further appraisal. Pathways representing primarily interventions with a number of published evaluations showing little effect, no effect or negative impact were removed as they were considered to have no explanatory power for the decrease in pregnancies. This produced the updated logic model with all suggested pathways from literature and systematic reviews.

## Analysis 2 – contribution to the UK’s decrease

### Selecting pathways for analysis

Using the logic model produced in Analysis 1, I selected the top candidate explanations for the UK’s observed fall in teen pregnancy rates from 1999 onwards. These were selected based on:

* The strength of evidence, if presented in the literature, of a large effect size on exposed subjects;
* The plausibility of the mechanism of change if evidence of effectiveness was scarce or absent;
* The applicability of the intervention or exposure to the UK.

### Evaluating likelihood of pathways

To test each of the selected pathways, I then conducted more specified literature searches in academic and grey literature for evidence of:

* Reach – Exposure to the intervention of all females under-18 in the UK, or a sizeable (targeted or untargeted) subsection, corresponding in time to observed drops in rates.
* Efficacy or Effectiveness – Efficacy is taken as a measure of the “success rate if implemented as in guidelines” (Glasgow et al., 1999, Table 1) for interventions, or suggested effects of a non-intervention pathway analysed by exposure of individuals to it. Effectiveness (“Efficacy × Implementation”; Glasgow et al., 1999, p. 1323) is taken in both cases as population-level assessment of the pathway’s effects on pregnancy rates without correcting for measures of implementation.
* Adoption – In cases of interventions, this was assessed by reports of relevant bodies’ uptake of intervention.
* Implementation – To what extent the “program is delivered as intended” (Glasgow et al., 1999, p. 1323). In cases of non-intervention exposures, this was taken as a measure of the difference between the strongest exposure observed in individual-level analyses and the expected or observed average level of exposure across the population.
* Maintenance – The extent to which the intervention or exposure is sustained over the relevant time period. The time period of interest is from 1999 to the present.

### Literature search

For each intervention, I searched for evaluation documents of the intervention and government and health-board reports of roll-out. I searched the academic literature for evaluations of interventions in the UK and supplemented this with relevant reference documents from the literature reviewed in Analysis 1.

For non-intervention exposures, I extracted relevant evidence for assessing a UK-wide exposure from the reviewed literature, then conducted further literature searches for other analyses of the hypothesised causal pathway.

Literature was screened for relevance to answering the questions presented in the RE-AIM framework for analysing programme effectiveness.

### Data analysis

To assess the evidence for an effect, the following questions were used, as adapted from the RE-AIM framework (Glasgow et al., 1999; Holtrop, Rabin, & Glasgow, 2018):

1. Reach – What proportion of under-18 females in the UK were exposed to the intervention? An intervention may have targeted a sub-population, however our focus is assessing its contributions to changing rates amongst the whole at-risk population.
2. Efficacy/Effectiveness – How did the intervention effect pregnancy rates amongst the exposed population when implemented per protocol (efficacy)? What were the effects of the intervention in real-world observations, whilst accounting for other factors (effectiveness)?
3. Adoption – What proportion of settings or institutions implemented this intervention within the policy/exposed context?
4. Implementation – To what extent was the intervention implemented as intended in the real world? [capturing dilution of protocols, cutting of funding, scarcity of resources, push-back, lack of training, etc.]
5. Maintenance – was the intervention maintained over the intended time scale? Did it continue to latest recorded observations of falling rates (2016)?

Each intervention or exposure was queried across all five domains. I analysed the presence of evidence available for answering each of the questions, the conclusions pointed to by the evidence and the quality of evidence presented. Combining this, I presented a narrative synthesis of the confidence of causal inference for each hypothesised pathway. I additionally highlight some areas requiring further or more reliable research to answer questions of causality or contribution to observed declines in pregnancy rates.

# Results

Talking points.

* 591 articles from literature searches. 39 additional articles from hand searching.
* 97 relevant reviews identified
* 5 of these were reviews of reviews (RoR) – I added relevant reviews from citations but did not include RoR data in analysis (vote duplication)
* All study conclusions were rated either + (weak evidence of positive effect) or 0 (no evidence of effect. No reviews found a consistent negative effect of an intervention
* In most +ve cases (in 91 reviews), evidence of some effect was presented in case-control or RCT studies of sample populations, but no assessment of population reach of an implemented intervention (Q4).
* In one case, exposure was measured across a population but observational data did not permit confidence in inferring causation (not controlled for counfounding, or presenting a logic model, or observed across contexts; Q1-3).
* Reviews reported on 30 interventions or environment changes associated with pregnancy; four environment changes which were previously highlighted for querying were not considered in the included reviews.
* The 34 interventions and environment changes were categorised into 19 distinct branches to be assessed for inclusion/exclusion within causal tree.

## Existing branches with supporting evidence:

### School-based sex and relationships education (SRE)

A large number of reviews (32) looked at school-based sex and relationships education programmes. In extraction, these were divided into pregnancy-focussed programmes (where avoiding pregnancy was a stated goal, alongside associated sexual risk behaviours) and STI-focussed programmes (where STI prevention was stated as an alternate or additional goal to pregnancy, and this was highlighted to participants). The majority of reviews (71%) examining evidence for STI programmes concluded a positive effect on pregnancy or pregnancy-risk behaviours, wheras a smaller majority (56%) found no or conflicting evidence from pregnancy-focussed programmes. Whilst some programmes may have delivered positive change, the No reviews presented evidence that school-based inerventions had contributed to population-level changes in teenage pregnancy rates.

There were overlapping aims across programme types (e.g. postponing sexual involvement, reducing number of partners, reducing frequency of sex), but potentially differences in several key elements with impact on pregnancy risk. STI programmes are likely to focus on condom use to prevent transmission whereas pregnancy programmes may recomment other contraceptives with less user-error risk. STI programmes motivate action through raising awareness of infection, whereas pregnancy programmes inform about pregnancy and childbearing, which may be received as a less alarming outcome.

### Abstinence-based education

Of the 33 revews which looked at abstinence-based (combining abstinence-only and abstinence-plus) interventions, 23 (70%) found no consistent impact, positive or negative, on teenage pregnancy rates.