



Feasibility study into evaluating the labour and childcare market impacts of Tax-Free Childcare and the Free Early Education Entitlement

**A REPORT PREPARED FOR HM REVENUE AND CUSTOMS AND
THE DEPARTMENT FOR EDUCATION**

HMRC Research Report 406

Disclaimer

The views in this report are the authors' own and do not necessarily reflect those of HM Revenue & Customs or the Department for Education.

© Crown Copyright 2016

Copyright in the typographical arrangement and design rests with the Crown. This publication may be reported free of charge in any format or medium provided that it is reproduced accurately and not used in a misleading context. The material must be acknowledged as Crown copyright with the title and source of the publication specified.

Published by HM Revenue and Customs, February 2016

www.hmrc.gov.uk

Authors and acknowledgements

Dr Gillian Paull is a Senior Associate and Ian Hathaway and Andrew Leicester are Managers in the Public Policy Practice at Frontier Economics.

The authors would like to thank Professor Mike Brewer of the University of Essex and Professor Eva Lloyd of the University of East London for valuable comments on the initial findings presented in this report. They would also like to thank participants at presentations at HMRC and the Department of Education for useful feedback and participants at an initial discussion with relevant stakeholders on the considerations for evaluation design.

Glossary

CCTC	Childcare element of WTC reimbursing 70 percent of childcare costs for parents in receipt of WTC (see WTC).
DiD (DiDiD)	Difference-in-differences (difference-in-difference-differences): evaluation methodology comparing changes in outcomes for the treatment group with those for a comparison group (and compared to another period or group without any treatment)
Dependent child	A dependent child is defined as one aged under 16 or aged 16-18 and in school or non-advanced further education, not married and living with a parent.
DfE	Department for Education
DWP	Department for Work and Pensions
ESC	Employer Supported Childcare: existing scheme for employees to receive a tax exemption and National Insurance disregard on childcare support provided by their employer.
EEEE	Free Early Education Entitlement: free places of 15 hours per week for 38 weeks a year for three and four year old preschool children and two year old children from the 40 percent most disadvantaged families. Planned extension to 30 hours per week for three and four year old preschool children of working parents.
Formal childcare	Registered, paid childcare including private, voluntary and independent providers, nursery schools, nursery classes in schools, Children's Centres and LA-maintained providers, childminders and nannies.
FRS	Family Resources Survey
HMRC	HM Revenue and Customs
Informal childcare	Unregistered (and often unpaid) sources of childcare including grandparents, other relatives and friends.

Intention to treat impact	Impact on all individuals who could alter their behaviour to become eligible for a policy and could be ‘treated’ rather than just those who take-up a policy.
LFS	Labour Force Survey
MDE	Minimum detectable effect: statistical estimate of the size of impact that can be identified as statistically significant for a given sample size.
ONS	Office of National Statistics
Parents	Adults with own, adopted or step dependent children living in the same household (see dependent child).
Providers	Providers of early education and childcare. For this study, this includes only formal childcare and excludes informal childcare (see formal and informal childcare).
RDD	Random discontinuity design: evaluation methodology using an arbitrary cutoff in eligibility to identify a treatment and comparison group.
TFC	Tax Free Childcare: planned scheme to pay 20 percent of childcare costs for children under the age of 12 for working parents not in receipt of tax credits or Universal Credit.
Triple difference	Alternative name for DiDiD (see DiDiD).
UC	Universal Credit: new scheme of government support being rolled out to replace a collection of existing benefits and tax credits including Jobseeker’s Allowance, Housing Benefit, Working Tax Credit, Child Tax Credit, Employment and Support Allowance and Income Support.
US	Understanding Society survey.
WTC	Working Tax Credit: existing scheme of means-tested income support for working individuals and families with children.

Feasibility study into evaluating the labour and childcare market impacts of Tax-Free Childcare and the Free Early Education Entitlement

Authors and acknowledgements	ii
Glossary	iii
Executive Summary	1
1 Introduction	6
2 Review of Existing Evidence	9
2.1 Methodologies.....	9
2.2 Sources of comparison groups	14
2.3 Impact metrics.....	15
2.4 Findings on impact sizes.....	16
3 Potential impacts	18
3.1 Overview of TFC and the FEEE.....	18
3.2 FEEE provision	20
3.3 Impacts on parental work and childcare choices.....	21
3.4 Secondary impacts on the labour market.....	24
3.5 Secondary impacts on the childcare market	24
3.6 Offsetting effects of childcare price rises.....	25
3.7 Summary of potential impacts	26
4 Overarching considerations for evaluation approaches	28
4.1 Identification of a comparison group	28
4.2 Concurrent other policy developments.....	29
4.3 Concurrent other influences on outcomes.....	30
4.4 Potential size of impacts	30
4.5 Timing of impacts.....	31

4.6	<i>Criteria for assessment</i>	31
5	Evaluation of employment impacts	33
5.1	<i>Impact metrics</i>	33
5.2	<i>Data sources</i>	34
5.3	<i>Options for TFC</i>	36
5.4	<i>Options for the FEEE 30 hour extension</i>	45
5.5	<i>Options for the FEEE previous extensions</i>	52
5.6	<i>Role of qualitative evidence</i>	53
5.7	<i>Supplementary structural modelling</i>	54
5.8	<i>Recommendations</i>	55
6	Evaluation of childcare market impacts	58
6.1	<i>Impact metrics</i>	58
6.2	<i>Data sources</i>	59
6.3	<i>Options for TFC and the FEEE 30 hour extension</i>	60
6.4	<i>Options for FEEE: previous extensions</i>	63
6.5	<i>Role of qualitative evidence</i>	63
6.6	<i>Recommendations</i>	64
7	Conclusion	66
	Annex A: Literature review of childcare policy evaluations	67
	Annex B: UK data sources	74
	Annex C: Power calculations	79
	References	84

Feasibility study into evaluating the labour and childcare market impacts of Tax-Free Childcare and the Free Early Education Entitlement

Figure 1. Logic Model for Tax-Free Childcare	19
Figure 2. Logic Model for the Free Early Education Entitlement	19
Table 1. Impact metrics used in existing literature	16
Table 2. Previous evidence on sizes of impacts	17
Table 3. Summary of potential impacts	27
Table 4. Impact metrics for parental employment	33
Table 5. Existing data sources for the employment evaluation	35
Table 6. Impact metrics for the childcare market	58
Table 7. Existing data sources for the childcare market evaluation	60
Table 8. Summary of evaluations: UK studies	67
Table 9. Summary of evaluations: other countries	68
Table 10. UK data sources: labour market / childcare usage	74
Table 11. UK data sources: childcare markets (providers)	75
Table 12. UK data sources: childcare markets (business data)	77
Table 13. Impact metric parameters used in the power calculations	80
Table 14. Sample sizes used in the power calculations	81
Table 15. Estimated mean detectable effects	82

Executive Summary

Introduction

Childcare policy in the UK has developed rapidly over the last two decades. Two new initiatives will be introduced in 2017 with the objective of supporting parents to work: Tax-Free Childcare (TFC) will offer a 20 percent subsidy for childcare costs to working parents and the Free Early Education Entitlement (FEEE) will be extended from 15 to 30 weekly hours for three and four year old pre-school children of working parents.

The aim of this feasibility study is to help HM Revenue and Customs (HMRC) and the Department for Education (DfE) develop an economic impact evaluation framework for TFC and the FEEE (both the previous extensions and the forthcoming 30 hour extension) to measure impacts on parental labour market outcomes and the childcare market. The first of these impacts captures the intended objective to raise parental employment by supporting parents to work if they choose to, while the second focuses on the potential impacts on childcare prices and the availability of places. It should be noted that the scope of the study does not cover the impacts on child development because there is an ongoing evaluation of the impacts of FEEE on child development being undertaken in the Study of the Early Education and Development (SEED).

Existing evidence

A review of the existing literature on evaluations of childcare policy highlights the importance and challenges of identifying a comparison group to estimate the counterfactual outcomes (what would have happened in the absence of the policy) and thereby estimate the size of impact. Indeed, evaluation of policy in the UK has been limited by the fact that policy changes have tended to have been introduced nationally; to be potentially available to all children of a particular age; and to have been introduced at overlapping times. Whilst the existing evidence suggests that childcare subsidies and free places can have small but significant impacts on parents' employment and childcare use, these can be difficult to evaluate. There is, currently, very little evidence on the impacts on the childcare market.

Potential impacts

Logic models tracking through the potential impacts of TFC and the FEEE indicate that, on balance, these policies create incentives to raise parental employment rates and average weekly work hours. However, the models also show that responses in the childcare market are very uncertain. On the one hand, TFC could raise the number of places, weekly hours and quality of care offered by childcare providers in response to higher levels of demand. On the other hand, it could simply increase prices and profits to a level which offsets the

intended impacts on parental employment. The FEEE 30 hour extension could lead providers to adjust many aspects of their childcare provision. It could impact on the pricing structure for all children and on the offer of places to children not of the age eligible for FEEE. Moreover, additional funding for FEEE 30 hour places could create pressures for prices to rise similar to those for TFC.

Challenges in evaluation design

Identifying evaluation approaches for the two policies faces several specific challenges:

- The concurrent introduction of TFC and the FEEE 30 hour extension together with the fact that the FEEE extension can have spillover effects to other ages of children makes it problematic to identify separate childcare market impacts for the two policies.
- A lack of consistent and robust data on childcare market metrics over a long period from which it would be possible to identify unusual changes with the introduction of new policies.
- The potentially small impacts on some outcome measures which mean that particularly large samples of data may be required to identify impacts.
- There could be some delay in impacts materialising after the introduction of the policies due to the time required for parent2 to obtain new work and childcare arrangements and for childcare providers to adjust their provision to new funding options or demand conditions. This means that evaluations may require further, later analysis if impacts are not identified in the shorter term.

Evaluation of TFC

1. Impacts on parental employment.

The evaluation of impacts on parental employment is feasible and likely to generate conclusive findings.

- The recommended approach is to use a difference-in-difference methodology with a comparison group of parents of dependent children aged over 11 applied to Labour Force Survey (LFS) data.

This approach has a high probability of identifying conclusive findings, can be undertaken at reasonably low cost and, based on current plans for the introduction and roll-out of the policy, could produce findings by 2019 (or 2020 if impacts take longer to materialise). The feasibility of this evaluation is driven by two factors. First, the existence of a comparable group of parents who face a similar employment context but are not eligible for TFC because their children are too old. Second, the availability of a large and robust data source with the required employment metrics.

Executive Summary

2. Impacts on childcare use.

The evaluation of impacts on childcare choices by parents is not a direct part of the remit of this study, but has been useful to consider. However, evaluating impact on childcare use will be more challenging than evaluating its effect on parental employment. Parents with only older children (over age 11) are a reasonable comparison group for employment outcomes, but are not a viable comparison group for childcare usage because their children are beyond the age when they generally require childcare. The absence of a good comparison group means that any evaluation of the impact of on childcare use will carry a risk of failing to generate conclusive findings.

- The recommended approach is to (optionally) consider the evaluation of the ‘treatment’ impacts on childcare use using a matched panel methodology based on TFC take-up applied to Understanding Society (US) data.

Although this approach carries a substantial risk that variation in TFC take-up may be insufficient to identify conclusive findings, the relatively low cost of the analysis of US data means the approach is worth considering.

3. Impacts on the childcare market

The options to evaluate the impacts on the childcare market are severely limited by the lack of a comparison group for the policy; by the concurrent introduction of the FEEE 30 hour extension; and by the absence of robust, consistently collected data on the childcare market impact metrics.

- The recommended approach is a simple discontinuity analysis of time trends using primary data collection to identify whether the introduction of both TFC and the FEEE 30 hour extension coincides with a distinct break in ongoing trends which cannot be attributed to any other factors.

However, this approach has substantial drawbacks including the high cost of primary data collection and the fact that the findings, while potentially consistent with hypotheses of joint impacts of both policies, would not provide a robust evaluation of impact for TFC alone or for the joint impact of both policies.

Evaluation of the FEEE 30 hour extension

4. Impacts on parental employment:

As with TFC, the evaluation of impacts of the FEEE 30 hour extension on parental employment is feasible and likely to generate conclusive findings.

- Similar to that for TFC, the recommended approach is a difference-in-difference methodology with comparison groups of parents of children aged 1-2 and parents of children aged 5-7 applied to LFS data.

This approach has a reasonable probability of identifying conclusive findings, can be undertaken at reasonably low cost and, based on current plans for the introduction and roll-out of the policy, could produce findings by 2019 (or 2020 if impacts take longer to materialise).

The similarity in methodology and data requirements for both the TFC and FEEE extension (and the need to consider the other policy) recommends that the evaluations of the impacts on parental employment would ideally be jointly undertaken.

5. Impacts on childcare use.

Evaluation of the impact on childcare use for the FEEE 30 hour extension can be undertaken with a reasonable likelihood that conclusive findings will be obtained.

- The recommended approach is to use, again, a difference-in-difference approach based on comparison groups of parents of children aged 1-2 and parents of children aged 5-7 applied to Family Resources Survey (FRS) data.

This approach has a reasonable probability of identifying conclusive findings, can be undertaken at reasonably low cost and, based on current plans for the introduction and roll-out of both policies, could produce findings by 2019/2020 (or 2020/2021 if impacts take longer to materialise).

6. Impacts on the childcare market.

An evaluation of the impact of the FEEE 30 hour extension on the childcare market faces identical challenges as evaluating the impact of TFC and can only be evaluated jointly with TFC.

- The recommended approach is therefore the joint analysis with TFC described in bullet (3) and carries the same substantial drawbacks, including that it would not provide a robust evaluation of impact for FEEE 30 hour extension alone or for the joint impact of both policies.

Retrospective evaluation of FEEE

The main findings for the retrospective evaluations of the extensions of the FEEE to three year olds from 2004 and to disadvantaged two year olds in 2013/2014 are as follows.

7. Impacts on parental employment and childcare use.

- The recommended approach is for DfE to continue to monitor the findings from ongoing research considering the evaluation of the employment and childcare use impacts of these policies using LFS and FRS data and data from the Study of Early Education and Development (SEED).

Executive Summary

8. Impacts on the childcare market

- An evaluation is not feasible due to a lack of adequate data sources.

Additional evidence

The quantitative evaluations could be supplemented with further information from three other sources:

- The potential impacts on parental employment and childcare use could be estimated prior to the introduction of the policies using simulations from a structural modelling approach. However, such estimates do not measure the actual impacts of the policies and are not a substitute for an evaluation. In addition, there are systematic biases which mean that these estimates may overstate the impacts and could, therefore, be regarded as ‘upper bound’ estimates of impacts.
- Qualitative research could usefully provide understanding of the mechanisms driving and reasons for the size of impacts identified in the quantitative evaluations and also inform on potential improvements in policy design.
- Qualitative research could also be a useful addition to the analysis of impact of the FEEE 30 hour extension on the childcare market: interviews with providers could potentially inform on how childcare provision had been affected by the FEEE 30 hour extension, independent of the impact of TFC.

1 Introduction

Childcare policy in the UK has developed rapidly over the last two decades, primarily beginning in the late 1990s with the introduction for Free Early Education Entitlement (FEEE) for four year olds and a substantial subsidy for childcare costs for working parents in the Working Families Tax Credit. In 2017, two further major developments are planned with the introduction of Tax-Free Childcare (TFC) and an extension of the FEEE to 30 hours for three and four year old pre-school children with working parents. The primary purpose of both of these policy developments is to support parents to work if they choose to.

The aim of this feasibility study is to help HM Revenue and Customs (HMRC) and the Department for Education (DfE) develop an economic impact evaluation framework for TFC and the FEEE 30 hour extension to measure impacts on two sets of outcomes:

- parental labour market outcomes and
- the childcare market

The first of these captures the intended objective to raise parental employment, while the second focuses on the potential impacts on childcare prices and the availability of places.

In addition, the study reviews the feasibility of similar retrospective evaluations for the extension of FEEE to three year olds from 2004 and to two year olds in the most disadvantaged families in 2013 and 2014.

It should be noted that two areas are not within the scope of this feasibility study:

- The policy scope of the study does not cover the impacts of the childcare element of Working Tax Credit or Universal Credit and therefore does not consider the impact of all sources of government support for childcare. An evaluation of Universal Credit is planned.
- The impact scope of the study does not cover the impacts on child development because there is an ongoing evaluation of the impacts of FEEE on child development being undertaken in the Study of the Early Education and Development (SEED)¹.

The introduction of TFC will, in part, replace the existing offer of Employer-Supported Childcare (ESC). ESC currently offers support for childcare costs to parents who work for an employer offering a scheme by providing a tax exemption and National Insurance contributions (NICs) disregard on childcare provision. This means that a parent can save approximately one third of their

¹ Further information on the evaluation can be found at <http://www.seed.natcen.ac.uk/>.

costs, up to £933 per year (though this varies depending on income level). Both parents can claim ESC, if their respective employers offer it. The support provided through ESC is only applicable for costs of using formal forms of childcare and the available amount is independent of the number of children using formal childcare. TFC will replace this scheme with a new mechanism of payment which is independent of whether the employer offers a scheme and also open to the self-employed, greatly increasing the number of potentially eligible parents. In comparison to ESC, TFC has a lower subsidy rate of 20 percent, but higher maximum amounts of £2,000 for each child per year (payable on a total expenditure of £10,000 for each child). Eligibility for TFC will require that all parents in the household expect to earn at least the equivalent of 8 hours wages at the national minimum wage per week in the coming quarter and that the family is not in receipt of tax credits or Universal Credit (effectively has earnings above the threshold to receive these credits). There may be an option for parents currently enrolled in ESC to remain in that scheme rather than switch to TFC². Given the low proportion of employers who currently offer ESC and the low proportion of parents who benefit from it, the introduction of TFC will effectively be a new option of support for childcare costs for most working parents. TFC will be introduced from early 2017.

The current FEEE offer in England³ is a free place for all three and four year old pre-school children and the 40 percent most disadvantaged two year old children of up to 15 hours per week for 38 weeks a year (with allowance for the offer to be spread over more weeks each year). The FEEE extension will offer an additional 15 hours per week for 38 weeks for all three and four year old pre-school children with the same work requirement as that for TFC (that all parents in the household expect to earn at least the equivalent of 8 hours wages at the national minimum wage per week in the coming quarter). Hence, it seeks to encourage parental employment both through reducing the cost of childcare for longer hours and through the work requirement. Further details are to follow in due course on what flexibility will be permitted in the use of the hours spread over more weeks; on whether funding rates for the 30 hour places will be permitted to vary from those for 15 hours places; and on requirements on providers to offer both the shorter and longer hour places. Full roll-out of the FEEE 30 hour extension will be from September 2017, with early implementation in some areas from September 2016.

In assessing the feasibility of evaluating the impacts of TFC and the FEEE 30 hour extension on parental employment and the childcare market, this study utilises several components. First, it reviews the existing literature on evaluations

² In addition, the workplace nurseries element of ESC will continue.

³ While ESC and TFC are UK-wide policies, FEEE is a devolved policy in Scotland, Wales and Northern Ireland. The scope of this study is only the FEEE policy in England.

of childcare policy to inform on possible methodologies and outcome metrics that might be used and on the potential size of impacts that might be expected (section 2). Second, it develops logic models of impacts on parental employment and the childcare market for each policy development, seeking to identify the mechanisms which map policies to outcomes; appropriate impact metrics and other influences which need to be considered (section 3). Third, it presents some overarching issues in the consideration of the evaluation options and the criteria by which different options can be assessed (section 4). Sections 5 and 6 then present the evaluation options for impacts on parental employment and the childcare market respectively. Each begins with a review of the appropriate impact metrics and relevant existing UK data sources before going on to present and assess the evaluation options and any supplementary analysis that could be useful. The final section of the report concludes.

2 Review of Existing Evidence

The review of the existing literature on the evaluation of childcare policy covered both UK and international research and had no set timeframe. This section summarises the findings from this review on the methodological approaches applied in previous work; the sources of comparison groups which have enabled evaluation; and the impact metrics that have been previously analysed. The final subsection provides some indication of the potential size of impacts for TFC and the FEEE 30 hour extension using the most appropriate available evidence.

Annex A provides further details on the evaluation studies considered and referenced in this section.

2.1 Methodologies

The review of the existing literature highlighted six types of evaluation approaches which have been used.⁴ Because these approaches overlap in some aspects, distinctions between them are not always clear-cut and actual applications can contain a mixture of methodologies. However, all approaches essentially seek a means to identifying a counterfactual with which to estimate what would have happened in the absence of the policy.

2.1.1 Random control trial (RCT)

Randomised control trials use deliberate random selection into a treatment group (eligible for the policy) and control group (not eligible for the policy). The random selection means that the two groups are equivalent in all respects except the policy, providing a perfect counterfactual. The main drawbacks of the RCT approach is that such trials can be expensive and difficult to implement, while the impacts cannot be generalised beyond the specific policy evaluated and there is no evidence on understanding the mechanisms which drive impacts.

There are relatively few examples of the use of RCTs involving childcare policy. These include:

- Welfare reform demonstration projects in the US, but few included specific elements involving the cost or provision of childcare.
- The Employment, Retention and Advancement (ERA) demonstration project in UK, but this did not separately identify the impact of the childcare element.

⁴ A previous review of methodological approaches to identify the links between maternal employment and childcare use is provided in Brewer & Paull (2004).

This type of approach is not considered for the evaluations of TFC or the FEEE 30 hour extension because it would have needed to have been included as part of the policy designs.

2.1.2 Non-randomised policy evaluations

Non-randomised policy evaluations can include area specific pilots or policies offered only to a subgroup of the potentially eligible population which means that counterfactual comparisons can be drawn with those areas or cases not in the pilot. The main drawback of this approach is that small sample sizes in the pilot area or group may be insufficient to identify quantitative impacts or the pilot may not last long enough to identify longer term impacts.

There are relatively few examples of the use of this approach involving childcare policy, but these include:

- The Neighbourhood Nurseries Initiative
- The Early Education Pilot for Two Year Old Children.

This approach is considered for TFC using the staggered introduction by age of youngest child. Use of the early implementation areas was also considered for the evaluation of FEEE 30 hour extension, but the number likely to be involved in the early stage is too small for a quantitative evaluation approach.

2.1.3 Statistical matching / regression analysis

This approach uses matching or multivariate regression to create a counterfactual for the treatment group who are subject to the policy.⁵ Matching reweights the comparison group to match the treatment groups in characteristics which affect the policy outcomes (e.g. propensity score matching), while multivariate regression analysis includes ‘control’ variables to ensure that the effects of other differences between the groups are not spuriously included in the estimated impact of the policy. The advantage of this approach is that this does not require a clearly defined comparison group of those not subject to the policy. The main drawback of the approach is the need for a rich dataset to ensure that matching or control variables include all individual characteristics and local conditions which might influence the outcomes. There is also a need to ensure that the regressions are correctly specified with appropriate control variables included in a manner (such as a linear or a quadratic relationship) which matches the nature of their influence on the impact metric.

There are no UK examples of the use of this approach with respect to childcare policy, but examples from international literature include Farfan-Portet, Lorant &

⁵ This approach is analogous to a ‘one difference’ approach relative to the ‘difference-in-differences’ approach described below.

Petrella (2011) (Belgium); Gustafsson & Stafford (1992) (Sweden); Felfe, Lechner & Thiermann (2013) (Switzerland); Ryan et al (2011), Weinraub et al (2005), Marrufo et al (2003), Davis & Li (2005) and Davis et al (2009) (USA) (details are provided in Annex A).

The main caveat on using this approach for the TFC or FEEE 30 hour extension evaluations is the need for a rich dataset to ensure that that matching or control variables include all individual characteristics and local conditions which might influence the outcomes.

2.1.4 Regression discontinuity design (RDD)

A regression discontinuity design relies on an 'arbitrary' rule whereby eligibility for the policy is determined by an individual's position in an observed continuous variable with an eligibility threshold, for example, having income below a certain threshold or a particular age of child. The approach compares individuals close to either side of the cut-off with any step-like discontinuity in the outcome at the eligibility threshold indicating an impact. The strength of this approach is that it is akin to a randomised experiment for individuals around the threshold.

The main drawbacks of this approach are that it only identifies impacts near the threshold cut-off and the data requirements are high as the sample selects only those close to the threshold cut-offs. To be robust, this approach also requires:

- No other discontinuities influencing the outcome at the same cut-off thresholds, e.g. a cultural perception that a child becomes ready to attend pre-school when they reach age three. In other words, that those just before or below the cut-off threshold are sufficiently similar to those just after or above it to provide a robust comparison group.
- No selective manipulation of the eligibility rules e.g. reducing income to qualify.
- No substitution between marginal participants and marginal non-participants, e.g. selection to accept some children early and some late.
- No anticipatory changes in outcomes or substantial delays in response.

Examples of UK studies using this methodology include Brewer et al (2014) and Brewer & Crawford (2010). Brewer et al (2014) use the eligibility cut-off for the FEEE for three year olds to estimate the impact on mothers' employment rates. Brewer & Crawford (2010) use the eligibility cut-off for compulsory school entry at age four to estimate the impact of school entry on lone parents' employment and welfare receipt. Studies from other countries using this approach include Berlinski, Galiani & McEwan (2011) (Argentina); Goux & Maurin (2010) (France); Bauernschuster & Schlotter (2015) (Germany); Fitzpatrick (2010, 2012) and Gelbach (2002) (USA).

2.1.5 Difference in difference (DiD)

The difference-in-difference (DiD) approach compares the change in outcome before and after treatment (typically before and after policy introduction) with a comparison group who remain ineligible over the same period (typically in another geographic area). The approach can be combined with matching in the comparison group and the use of control variables which may also affect changes in the outcome variables. It can be applied using panel data (same individuals observed before and after) and with repeated cross-section data (different individuals observed before and after, but sampled from the same population). The main advantage of this approach is that it can control for all observed and unobserved differences between the treated group and the counterfactual.

Robust estimation using this approach requires several conditions:

- No differential trends between the treated and comparison areas (although absolute levels can differ). This can be tested or controlled for using a ‘triple difference’ (difference-in-differences-in-differences (DiDiD)) addition, whereby differential trends between the treated and comparison areas are tested against another group (e.g. different age of children) or during an earlier period when there is no change in policy. This is sometimes also called a ‘placebo’ test. The drawback of the triple difference approach is that it can reduce the precision of the estimated impact.
- That the decision to implement the policy or the decision of individuals to use a policy is not related to transitory shocks to the outcomes. This can be addressed by testing whether the policy can be explained by any outcomes prior to implementation.
- When using cross-section data, there must be no changes in the composition of the treatment or control group between the initial point and the point after the policy change.

Examples of UK studies using this approach include Blanden et al (2014) and Brewer et al (2014) who use the speed of implementation across geographical areas to evaluate the impact of the introduction of FEEE in England for three year olds in 2004 on child outcomes and maternal employment respectively. Studies from other countries using this approach include Berlinski & Galiani (2007) (Argentina); Baker, Gruber & Milligan (2008), Lefebvre & Merrigan (2008, Lefebvre, Merrigan & Verstraete (2011) (Canada); Bauernschuster & Schlotter (2015) (Germany); Schlosser (2011) (Israel); Bettendorf, Jorgen & Muller (2012) (Netherlands); Havnes & Mogstad (2011) (Norway); Nollenberger & Rodriguez-Planas (2015) (Spain); Lundin et al (2008) (Sweden); Bassok, Fitzpatrick & Loeb (2014), Cascio (2009) and Casio, Whitmore & Schanzenbach (2013) (USA).

Review of Existing Evidence

2.1.6 Simulations from structural models

The final approach is not, strictly speaking, an evaluation methodology but an alternative method to estimate the impact of a policy.⁶ Simulations of impacts using structural modelling require no actual policy to have been implemented. In the case of childcare policy, estimates of the responses of parents to differences in childcare price are used to predict the impacts of childcare subsidies or provision of free childcare. The requirements to undertake such modelling and the advantages and drawbacks are discussed in more detail in section 5.7 which considers how this approach might be used to help inform on the potential impacts of TFC and the FEEE 30 hour extension.

Examples of UK studies using this approach include Blundell et al (2000), Jenkins & Symons (2001), Paull, Taylor & Duncan (2002) and Viitanen (2005) who estimate the impact of childcare subsidies on childcare use and maternal employment. Studies from other countries include Doiron & Kalb (2005), Gong, Breunig & King (2010), Kalb & Lee (2008), Rammohan and Whelan (2007) (Australia); Cleveland, Gunderson & Hyatt (1996); Powell (1997, 1998, 2002) (Canada); Michalopoulos & Robins (2000, 2002) (Canada & USA); Chone et al (2003 (France); Coneus, Goeggel & Muehler (2009), Muller & Wrohlich (2014), Wrohlich (2011) (Germany); Oishi (2002) (Japan); Kornstad & Thoresen (2007) (Norway); Andren (2013), Brink, Nordblom & Wahlberg (2007) (Sweden); Averett et al (1997), Blau & Hagy (1998), Blau & Robins (1988, 1989), Blau & Tekin (2007), Connelly (1992), Kimmel (1998) and Ribar (1992, 1995) (USA).

2.1.7 Key considerations in evaluation design

The review of the existing literature identified the following key considerations in the design of an evaluation of childcare policy.

1. The key consideration is the identification of a counterfactual comparison group which is identical to the treatment groups or for which there is sufficient information for the evaluation to control for any differences. It is also important that the process determining where or when the policy was introduced (or speed of introduction) is not determined by any factors related to the outcome metrics (i.e. policy is not endogenous).
2. The sample size of the treatment and comparison groups must be sufficient to test the statistical significance of estimated impacts (e.g. identifying a sufficiently large sample of children close to the cut-off threshold for an RDD approach).
3. Legacy effects of ongoing policy (e.g. remaining eligibility for ESC in the evaluation of TFC) can have two potential effects. First, they could affect the

⁶ Much of the evidence cited in the Impact Assessment for Tax-Free Childcare is drawn from the studies using this approach (HM Revenue and Customs (2014)).

composition of the comparison and treatment groups over the period of implementation, leading to a bias in the estimated impact. Second, they could mean that the estimated impact is only applicable to the situation with the legacy ongoing and not necessarily to later situations when the legacy is no longer operating.

4. Initial conditions may affect the size of the estimated impact (e.g. the high proportion of three year old and four year old preschool children in early education could limit the size of potential impact on this metric). This could be particularly important when comparing impacts between different groups.
5. There is an important distinction between ‘intention to treat’ and actual treatment impacts (e.g. all parents with a child under the age of 12 are included in the intention to treat for TFC, but only a proportion are expected to take-up TFC). The first measures the impact on all individuals who could alter their behaviour to become eligible and be ‘treated’. This measure effectively spreads the average impact across all of the individuals that the policy might have intended to treat. The second measures the impact only on those individuals who take-up the policy and spreads the average impact across only those who the policy actually affected (if only in their take-up of the policy).
6. Anticipation effects can affect outcomes in the comparison group and/or the composition of the treatment and comparison groups, creating a biased estimate of the impact (e.g. parents may enrol their child in early education prior to the threshold for a free place in anticipation of the free place).
7. Delays in responses to the policy can be important for the timing of the data required for the evaluation. Allowance must be made for time for the policy to be implemented, for awareness of the policy to be established, for take-up, and for changes in behaviour (e.g. time to obtain work after the decision to work is made).

2.2 Sources of comparison groups

The literature review highlighted that good sources of comparison groups to evaluate childcare policy are relatively limited and that ongoing comparisons (in comparison to those based on policy introduction) are rare. The main sources have been:

- Geographical variation in the speed of implementation of the FEEE for 3 year olds in England; of the expansion in free pre-school in Argentina; of free pre-school for 3-4 year olds in Israel; of subsidised childcare for 3-6 year olds in Norway; and of subsidised childcare for 3 year olds in Spain.

Review of Existing Evidence

- Introduction of the \$5 a day childcare subsidy in Quebec but not elsewhere in Canada.
- Introduction of highly subsidised kindergarten for 3-6 year olds, but not other ages of children in Germany.
- Increase in childcare subsidies for families with younger children compared to those with only older children in the Netherlands.
- Geographical variation in childcare price reductions due to the introduction of a price cap in Sweden.
- Geographical variation in the provision of kindergarten (focused on Georgia and Oklahoma) in the USA.

Options to evaluate policy in the UK have been limited by three factors in particular. First, policy changes have tended to be introduced nationally (with the exception of devolved FEEE policies). Second, policies tend to be potentially available to all children of a particular age. Finally, policies have tended to be introduced at overlapping times.

2.3 Impact metrics

Table 1 presents a list of the impacts metrics that have been considered in evaluations of childcare policy.

The key metrics have generally been mothers' work participation and work hours (often divided into the discrete categories of part-time and full-time work); fathers' work participation and work hours; and the proportion of children using formal childcare. Other metrics have been more rarely used, typically only in a single study.

Variation in impacts across different groups have also been analysed for single and married mothers; for mothers with and without younger children (than targeted in the policy); for mothers of different levels of education and for different income groups. Very occasionally, evaluations have sought to distinguish between short run and longer term effects by analysing data from different durations after the policy introduction. For example, Lefebvre & Merrigan (2008) found that the size of the effects of the Canadian \$5 a day childcare subsidy on mothers' employment, hours worked, weeks worked and earnings increased over the five years following policy introduction. However, it is not clear whether this reflected a delay in implementation rather than impact as they also state "we observe the effect of the policy to be stronger as more subsidized spaces are offered to mothers of young children" (page 537).

Table 1. Impact metrics used in existing literature

Parental employment and childcare use metrics	Childcare market metrics
Work participation of mothers	Number of private and public childcare providers *
Whether mother works or studies *	Employment in private and public childcare providers *
Work hours for mothers	Pay per employee (as a proxy for quality)*
Part-time and full-time work for mothers	Quality of care *
Annual weeks worked for mothers *	Childcare price
Annual earnings for mothers *	
Work participation of fathers	
Work hours for fathers	
Welfare receipt for lone mothers *	
Proportion of children using formal childcare	
Weekly hours of formal childcare per hour *	
Amount spent per hour of childcare *	

Notes: * Only evaluated in a single study.

2.4 Findings on impact sizes

Table 2 presents some estimates of the sizes of impacts on the key parental employment and childcare use metrics, primarily from UK studies. These have been supplemented with evidence from studies in Australia for two of the metrics where UK evidence is not available. It should be noted that most of the estimates are drawn from evidence based on simulations using structural modelling rather than evaluations of actual policies. In addition, some of the impacts have been interpolated from those reported to provide a comparative measure at 20 percent and 50 percent subsidies. Finally, it should be noted that several of these studies were undertaken on data for some years ago when many of the impact metrics were lower and there was greater potential for the childcare policy to have a larger impact.

The existing evidence suggests that childcare subsidies and free places can have small but significant impacts on parents' employment or childcare use. These numbers also provide a basis from which to judge whether the evaluation approaches suggested below may have the capability to detect impacts of TFC or the FEEE 30 hour extension. However, there is currently very little evidence on the impacts on the childcare market.

Review of Existing Evidence

Table 2. Previous evidence on sizes of impacts

Impact metric	Evidence on the size of impacts
Mothers' employment rate	<p>20% / 100% childcare subsidy → 2 / 4 percentage point increase for lone parents (Jenkins & Symons (2001))</p> <p>20% / 50% childcare subsidy → 3 / 5.5 percentage point increase for married mothers (Viitamen (2005))</p> <p>Free 15 hours places for 3 year olds → 2 to 4 percentage point rise (Brewer et al (2004))</p>
Mothers' hours of work	<p>20% / 50% childcare subsidy → 0.4 / 1.1 hours increase for lone parents (Kalb & Lee (2008), Australia)</p> <p>20% / 50% childcare subsidy → 0.1 / 0.3 hours increase for mothers in couples (Kalb & Lee (2008), Australia)</p> <p>20% / 50% childcare subsidy → 2.4 / 6.0 increase for married mothers (Gong, Breunig & King (2010), Australia)</p>
Fathers' employment rate and hours of work	Childcare subsidy → no change in employment rate or hours (Kalb & Lee (2008), Australia)
Proportion of children in formal childcare	<p>20% / 50% childcare subsidy → 3.0 / 5.9 percentage point rise (Viitamen (2005))</p> <p>Childcare subsidy → rise for pre-school children, no impact for school age children (Paull et al (2002))</p> <p>100 free part-time places for three year olds → 14 children begin to use formal childcare (Blanden et al (2014))</p>
Average weekly hours of childcare per child	Childcare subsidy → no impact for pre-school or school age children (Paull et al (2002))
Amount spent per hour of childcare	Childcare subsidy → increase in amount spent per hour (Paull et al (2002))

Notes: Evidence is based on simulations from a structural modelling approach in Gong, Breunig & King (2010), Jenkins & Symons (2001), Kalb & Lee (2008), Paull et al (2002) and Viitamen (2005). Further details on the studies are presented in Annex A.

3 Potential impacts

In order to inform on the evaluation design, logic models for each policy were developed which map the mechanisms by which the policies might impact on parental employment and childcare market outcomes. These models can help to identify appropriate impact metrics (both target outcomes and other possible consequences) and other influences on policy impact which need to be considered in the evaluation design.

3.1 Overview of TFC and the FEEE

Figures 1 and 2 present the logic models for the potential impacts of TFC and the FEEE respectively.

Both models present impacts on parental work and childcare choices; on the decisions of childcare providers in their delivery of childcare; and on the labour market. There is relatively little flow in these logic models because many of the decisions are simultaneously interlinked and it is not often the case that one decision leads on to another. For example, parents choose between different potential packages of work and childcare choices rather than obtaining work and then considering childcare options (or vice versa). Similarly, providers must consider their childcare provision as a bundle of places and prices rather than setting one and then deciding upon the other.

The two figures highlight that the ordering of the mechanisms of impact are notably different between the two policies. TFC is essentially a demand-side driven childcare policy: the policy initially impacts on childcare demand from parents with potential subsequent reactions by childcare providers in their supply decisions. The FEEE is a supply-side driven childcare policy: the policy initially impacts on childcare providers in their decision to supply free places with potential subsequent reactions by parents in their use of childcare and work choices. For both policies, these subsequent impacts can have feedback responses which could potentially partially (or completely) offset the initial impacts.

In spite of the difference in ordering of impacts, the policies have many effects in common and the influences on the impacts are broadly similar. For this reason, the description below of the processes driving the impacts is organised by the different steps rather than independently for each policy.

Figure 1. Logic Model for Tax-Free Childcare

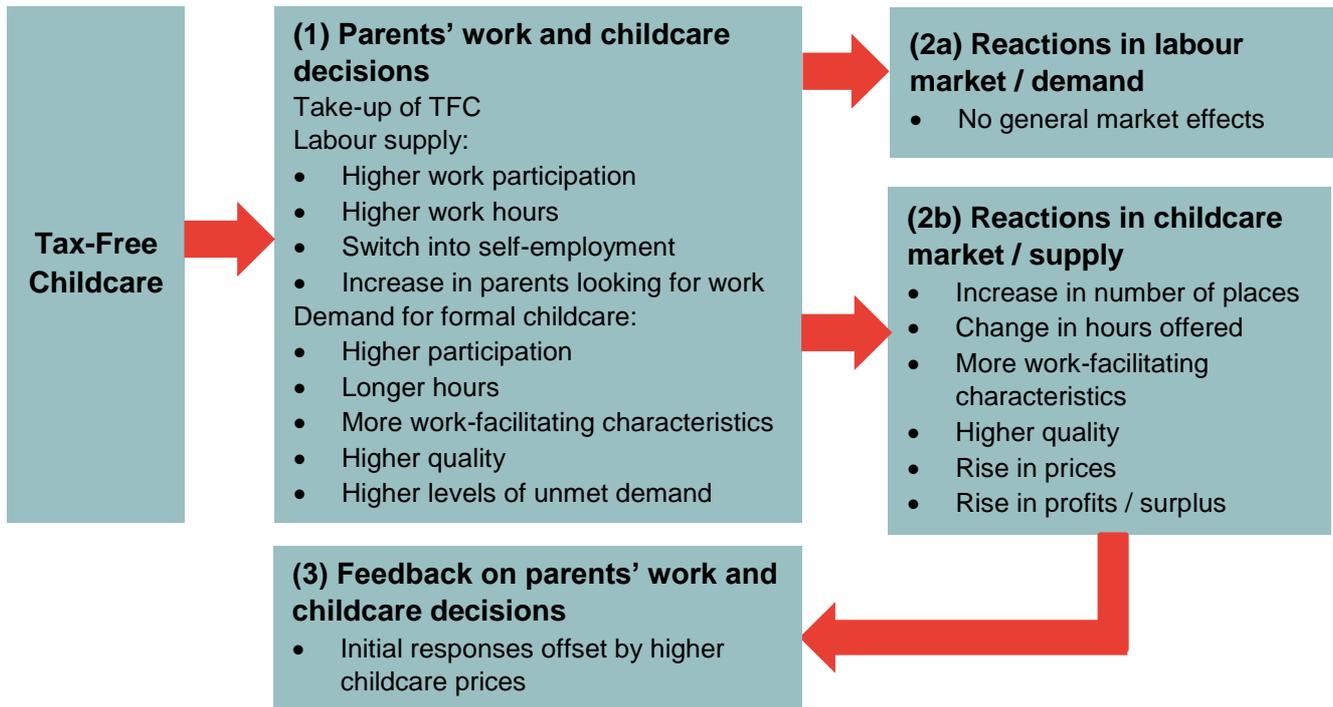
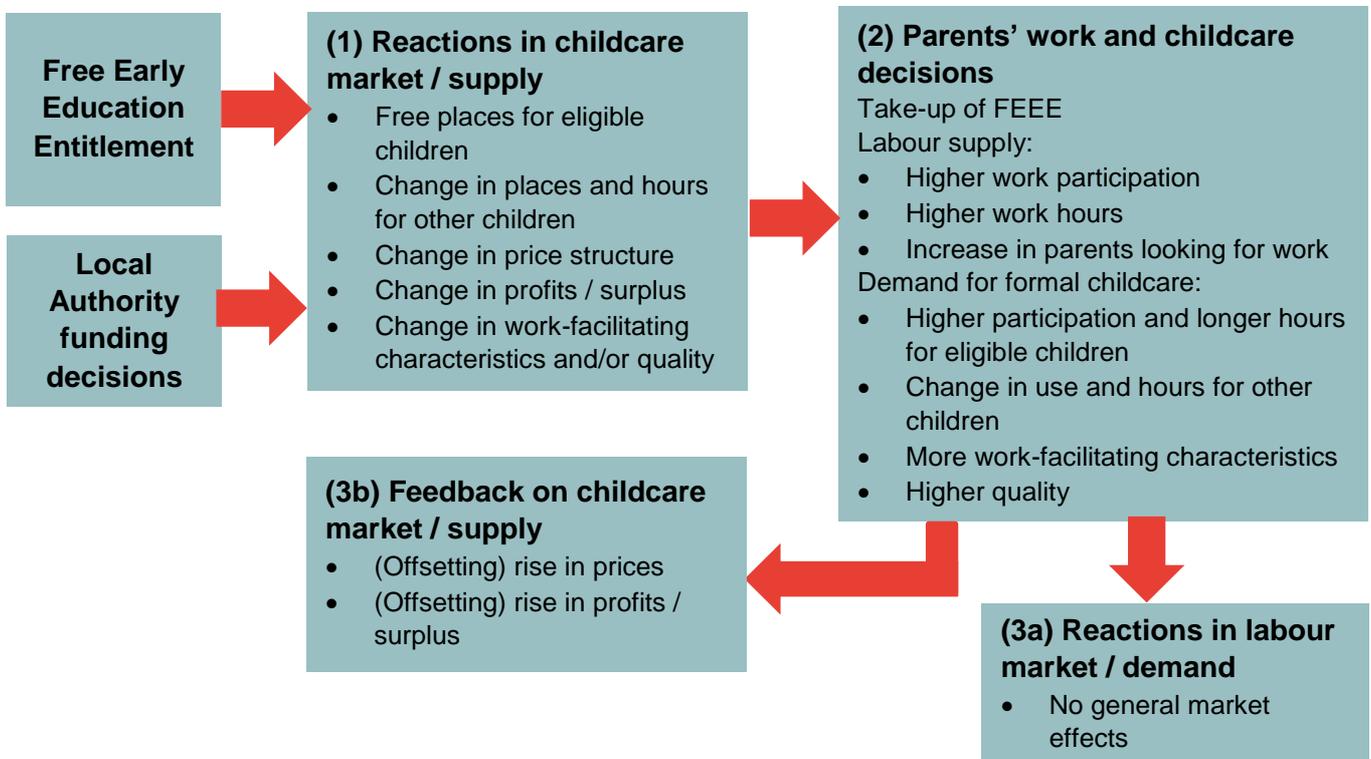


Figure 2. Logic Model for the Free Early Education Entitlement



3.2 FEEE provision

In order to consider the impacts on parental choices of TFC and the FEEE together, it is necessary to first consider how the FEEE impacts on the childcare provision offered to parents (summarised in box 1 in figure 2).

The FEEE has an immediate impact on provision:

- Increase in free places for eligible children.⁷
- Providers may adjust the number of places or hours offered for non-eligible children. For example, free places could either squeeze out or complement extension of other provision.
- Providers may adjust their prices for additional paid hours for children taking up free places and for non-eligible children, potentially allowing cross-subsidisation between funding for free places and parental fees for purchased hours.
- For private providers, the changes in provision and funding could affect profit levels. For non-private providers (such as the maintained sector and voluntary and independent providers), it may affect the degree to which they can extract a non-profit return ('surplus') such as being able to offer staff better conditions or pay.

There could also be an impact on the characteristics or quality of childcare offered:

- The change in provision and funding levels for free places could also involve changes in the characteristics of childcare, in particular between what can be termed 'work-facilitating characteristics' (such as childcare at atypical hours or for flexible hours) and 'higher quality' (typically used to mean childcare characteristics associated with better child development but used more broadly here to mean characteristics which are considered beneficial for the child). Dependent upon funding levels for the free places, providers may be able to offer more costly characteristics or higher quality or may have to offer lower cost options.

There are several factors which can influence the impact of the FEEE on childcare provision.

⁷ However, it should be noted that while the introduction of the 15-hour FEEE could only increase the number of free places, the 30-hour offer could potentially reduce the number of free places offered for two reasons. First, if there are constraints on the total hours that can be delivered, some providers may reduce the number of 15 hour places to offer 30 hour ones. Second, if providers are required to offer places at both hours levels (although this is not currently planned to be the case), some might be unwilling to offer 30 hour places and consequently cease offering the 15 hour places.

First, a critical factor influencing the impacts on childcare provision is the level of funding for free places set by Local Authorities (LAs) (highlighted in a separate box in figure 2). The extent to which funding covers the cost of provision will determine whether providers can afford to offer free places, particularly for the FEEE 30 hour extension which has less flexibility than the 15-hour FEEE to cross-subsidise costs with parental fees for additional purchased hours. Higher levels of funding may have an opposite effect, allowing providers to potentially reduce prices for purchased hours. In addition, the bureaucracy involved and the reliability of payment of FEEE funding could also influence whether providers decide to offer free places.

Second, the cost structure and resource constraints faced by providers will influence their ability to expand provision and offer more places or longer hours. For example, there could be constraints in obtaining additional venue space, investment financing or qualified staff. In particular, the hourly cost per child could be higher (or lower) for longer weekly hours with the FEEE 30 hour extension if average costs rise (or fall) with longer hours per child.

Finally, the competitiveness of the market and providers' ability to raise prices for purchased hours will influence how they alter their overall pricing structure in response to offering free places.

3.3 Impacts on parental work and childcare choices

Both policies have potential impacts on parents work and childcare choices (summarised in box 1 in figure 1 and box 2 in figure 2). These decisions are made jointly in conjunction with the decision on whether to take-up TFC or a free entitlement place.

3.3.1 Take-up of TFC / FEEE

An initial point to note is that take-up of TFC or FEEE can be high with little impact on parental work or childcare choices. In the case of TFC, while some families with a child under age 12 may have to alter their work or childcare choices to be eligible for TFC, families who already meet the work and childcare use requirements may similarly simply take-up TFC without any change in their work or childcare choices. In the case of the 15-hour FEEE, take-up of places may only involve a switch from a paid place to a free place with no effective change in childcare, while families already using longer hours of childcare and meeting the work requirements may take-up the FEEE 30 hour offer. If take-up of either policy does not involve substantial changes in work or childcare choices, the main impact of the policy is to effectively supplement the income of working families.

3.3.2 Impacts on parental work choices

TFC and FEEE have two key potential impacts on parental work choices:

- They increase the incentives for parents to be in work, both because of the work requirements (in TFC and the FEEE 30 hour extension) and because the reduction in the cost of childcare needed to facilitate work means that the financial return to working is greater.⁸
- They increase the incentive to work longer hours by reducing childcare costs and increasing the financial return to working.⁹

Hence, the primary expected impacts on parental employment are to raise work participation and increase work hours. However, an important influence on whether changes in parental work choices are realised is the extent to which suitable work opportunities are available for parents to enter work or extend their work hours.

There are also two secondary potential impacts on parental employment:

- Because the TFC entitlement will also be available to the self-employed whereas the ongoing policy of Employer-Supported Childcare is only available to the employed, it could encourage some parents to switch from employment to self-employment.
- If parents are unable to find suitable work opportunities, there may be a rise in the number of parents looking for work.

3.3.3 Impacts on parental childcare choices

The policies have similar primary potential impacts on childcare choices:

- Participation in formal childcare may rise for children eligible for TFC or the FEEE because childcare is effectively cheaper or, in the case of TFC and the FEEE 30 hour extension, because parents may need to use childcare to allow them to work in order to be eligible.¹⁰

⁸ There are also some more minor potential disincentives to participate in work in the 15-hour FEEE policy. First, the free places may reduce the need for earnings to pay for childcare used for non-work-related reasons. Second, because eligibility for free places for two year olds is based on having income below a threshold, it may create a disincentive to participate in work.

⁹ There are also more minor potential incentives to work fewer hours. First, the implicit income supplement from TFC or the FEEE means that the same level of living standards can be maintained with lower earnings and may reduce the number of hours that parents want to work. Second, the primary worker may reduce their work hours if the other parent in a couple starts working in response to TFC or the FEEE 30 hour extension. Empirically, people who can command higher wages are more likely to work and tend to work longer weekly hours, indicating that the positive impacts of the higher financial return to working is likely to outweigh the negative impact of the income subsidy.

¹⁰ This may involve a switch away from use of informal childcare (such as that provided by grandparents) which could potentially increase formal labour supply and employment among other (older) workers.

Potential impacts

- Weekly hours of formal childcare may also rise for eligible children because childcare is effectively cheaper or, in the case of TFC and the FEEE 30 hour extension, because parents need longer hours to facilitate working.¹¹
- There could be an increase in the total hourly amount spent on childcare (the combined subsidy and parental fees) for two reasons. First, parents may require more expensive childcare with characteristics which help facilitate working (such as flexible hours) in order to work or work longer hours for TFC or the FEEE 30 hour extension. Second, the implicit additional income from TFC or the FEEE may allow them to purchase more expensive care either with work facilitating features or with characteristics that they consider beneficial for the child. In the extreme, the entire TFC subsidy could be used to purchase more expensive childcare rather than any change in the amount of childcare used.

Hence, the primary expected impacts of both policies on childcare usage is to raise participation in formal childcare, the weekly hours of childcare and the hourly amount spent on childcare. However, the extent of these changes will be influenced by whether suitable childcare places or longer hours are available for parents to use and on parental willingness to use formal rather than informal types of childcare.

There are also two secondary potential impacts on parents' childcare choices:

- In the case of the FEEE, there could be impacts on the childcare choices for other children in the family not directly eligible for either policy for several reasons. First, because providers have altered the available places or prices for other children in conjunction with offering the free places. Second, because the implicit income supplement in the FEEE enables parents to spend more on childcare for other children or because use of free places for one child facilitates the use of childcare for other children. Third, increased work may raise the need for childcare for all children.
- In the case of TFC, if there is a rise in demand for formal childcare places or hours but parents are unable to find suitable childcare to meet that demand, there could potentially be an increase in reports of unmet demand for childcare.

¹¹ On the other hand, use of the 15-hour FEEE places could potentially reduce weekly hours of childcare if parents who would otherwise have used longer hours find it difficult to obtain additional hours to use with the 15-hour FEEE.

3.3.4 Cross constraints in parental choices

Finally, it should be noted that there are potential ‘cross constraints’ on parental work and childcare responses. For example, a lack of suitable childcare could prevent parents from entering work or a lack of employment opportunities could prevent parents increasing their use of formal childcare. The size of impacts on parental employment could be limited by either a lack of employment opportunities or suitable childcare places.

3.4 Secondary impacts on the labour market

Any change in parents’ labour supply could potentially have reactions in the labour market (summarised in box 2a in figure 1 and in box 3a in figure 2). However, parents are only a small fraction of all workers and it is extremely unlikely that changes in parental labour supply induced by TFC or the FEEE would have any overall labour market impacts in terms of aggregate wage levels or displacement of other workers. Potential impacts might occur in local, segmented markets (e.g. demand for part-time work in rural areas), but it can be assumed that, in general, there are no second round impacts on the labour market.

3.5 Secondary impacts on the childcare market

Changes in parental demand for childcare induced by TFC or the FEEE may have more substantial impacts on the childcare market (summarised in box 2b in figure 1 and box 3b in figure 2).

Increased parental demand for childcare places, for longer hours of childcare or for more expensive types of care in response to TFC or the FEEE could lead to two types of reactions in the supply of childcare:

- On the one hand, childcare providers could respond by delivering more places, longer hours and more expensive types of childcare demanded by parents.
- On the other hand, the increased demand could allow providers to raise prices and increase their profits or surplus.

Actual reactions in the childcare market may lie somewhere between these two extremes. However, it is difficult to predict the likely outcome for several reasons.

First, as in any market, reactions to increased demand will depend upon the degree of competitiveness in the market. If the market is less competitive, providers can push up prices to maximise profits without fear of losing customers and the outcome is more likely to be a general price rise rather than increased provision. But there is no direct evidence on the degree of

Potential impacts

competitiveness in the childcare market to predict the extent to which prices could rise.

Second, the childcare market in the UK is a ‘mixed market’ with a minority of provision delivered by the maintained sector (government providers such as nursery classes in schools, nursery schools, Children’s Centres and other Local Authority provision), and by voluntary and independent providers.¹² These types of providers may be less motivated to maximise profit and prefer not to raise prices in the face of increased demand. On the other hand, they may respond with price rises for non-profit reasons, such as, being able to offer staff better conditions or pay or being able to provide better ‘quality’ of care. It should also be noted that it is sometimes argued that private childcare providers are not primarily motivated by profit and may prefer not to raise prices in the face of excess demand. Without a clear understanding of providers’ motivations, it is challenging to assess whether and how much prices might rise.

Finally, providers’ reactions may be limited by the cost structure that they face. Increasing provision in the short run may only be possible with rising costs, for example, expansion may cost more per hour if staff must be paid higher overtime hourly wages or more expensive agency staff are used. In the longer term, resource constraints on expansion such as shortages in venue space; financing to expand or a lack of qualified staff could mean that expansion can only be achieved at higher costs and thereby prices. Again, evidence on cost structures and constraints on expansion is weak, making it difficult to predict the impact on prices.

3.6 Offsetting effects of childcare price rises

Any price rises in the childcare market will have a feedback effect on parental choices by offsetting the reduction in the effective childcare price brought about by TFC or the FEEE (shown in box 3 in figure 1 and box 3b in figure 2).

In the case of a universal childcare subsidy (without any age restrictions or work requirements), prices could rise up to a level equal to the subsidy. At such a level, parents would effectively face the same prices as in the absence of the subsidy and would not alter their work or childcare choices. The impact of the policy could be to simply raise providers’ profit or surplus.

The potential maximum impact of TFC and the FEEE on childcare prices is complicated by the fact that they only offer subsidies for a subset of childcare usage (only for children of working parents above an earnings threshold in the

¹² Out of the 107,900 providers of childcare in England in 2011, 65 per cent were private, 14 per cent voluntary, 2 per cent run by local authorities and 19 per cent were school or college based (derived from Brind et al. (2012a), tables 3.1 and 3.5a and 3.5b categorising childminders as private and early years as school or college based and including early years’ care in maintained schools).

case of TFC and only for children of the target age in the case of the FEEE). This also means that there are distributional consequences in that the price rises may affect some parents who not eligible for the subsidy or who choose not to take up the subsidy. Nevertheless, the key point remains that price rises could potentially offset any impacts on parental employment and childcare usage.

3.7 Summary of potential impacts

Table 3 provides a summary of the potential impacts. The main difference between TFC and the FEEE are the uncertain impacts in the childcare market for the FEEE, created by the uncertainty surrounding funding levels for the free places. In addition, the FEEE 30 hour extension might be expected to have greater impacts on parental employment than the 15 hour FEEE due to the work requirement and longer hours of childcare.

Table 3. Summary of potential impacts

Outcome	TFC	FEEE 15 hours	FEEE 30 hour extension
Parental employment			
Work participation	Increase	Increase	Increase
Weekly work hours	Increase	Increase	Increase
Self-employment	Increase	None	None
Number of parents looking for work	Increase	Increase	Increase
Formal childcare use			
Participation	Increase	Increase	Increase
Weekly hours	Increase	Increase	Increase
Hourly expenditure	Increase	Increase	Increase
Reports of unmet demand	Increase	None	None
Childcare market			
Number of free places	None	Increase	Increase
Number of paid places	Increase	Uncertain change	Uncertain change
Weekly hours per child	Increase	Uncertain change	Uncertain change
Prices	Increase	Uncertain change	Uncertain change
Profits / surplus	Increase	Uncertain change	Uncertain change
Quality of childcare	Increase	Uncertain change	Uncertain change
Work facilitating features	Increase	Uncertain change	Uncertain change

4 Overarching considerations for evaluation approaches

This chapter highlights some general issues and assumptions which apply to the consideration of evaluation design for all three policies and for both sets of impacts on parental employment and the childcare market.

4.1 Identification of a comparison group

A key challenge in any evaluation is to find a suitable comparison group to estimate counterfactual outcomes for the treated group which is subject to the policy. For the childcare policies considered here, three types of comparison groups can be ruled out:

- Families with children aged under 12 who claim Universal Credit (UC) or any tax credits are not eligible to claim TFC but they could potentially make alternative work and childcare choices which would mean they could claim TFC. Hence, patterns of work behaviour for UC and tax credits claimants may be affected by the TFC policy and these families cannot be used as a comparison.
- There is no potential for geographical variation in the speed of implementation of TFC which could be used in an evaluation of impacts.
- TFC is a UK-wide policy and there is no potential to use the devolved nations as a comparison group. Although FEEE is a policy specific to England, the use of the devolved nations as a comparison group is ruled out partly because the sample sizes are small and an evaluation using Wales and Scotland as a comparison group would only detect impacts if they were larger (see Annex C). Moreover, such an approach would also require that closely related policies in the devolved nations remain unchanged during the period when the FEEE 30 hour extension is introduced in England.

However, comparison groups with other ages of children will be considered in the evaluation of the FEEE. The logic models in the previous section highlighted that there can be spillover effects to children of non-eligible ages both through changes in parents' childcare choices for non-eligible children if they have a child receiving the FEEE and through providers' reactions in their provision and prices for non-eligible children if they offer the FEEE. It is assumed that these spillover effects are important for childcare market outcomes including the number of places and prices, but are of minor importance for parental work and childcare choices. The latter is assumed because the spillover effects only operate for families who have a child eligible for the FEEE or as secondary reactions to the adjustments made by providers in provision for non-eligible children. This

assumption means that comparisons with non-eligible ages of children can be used in the evaluation of the impact of the FEEE on parental employment but not in the evaluations of impacts on the childcare market.

4.2 Concurrent other policy developments

Concurrent other policy developments are a concern for the evaluation design, particularly if they potentially affect the group subject to the policy of interest and the comparison groups differently. There are two issues to consider in this respect.

First, Universal Credit (UC) is gradually being rolled out and the proportion of childcare costs that can be reimbursed under UC is planned to rise from 70 percent to 85 percent in spring 2016.¹³ Although the timing of this change does not overlap with either the introduction of TFC or the FEEE 30 hour extension, it will be important to ensure that any estimation of prior trends in parental employment and childcare market metrics are over a sufficiently long prior period to check whether trends were different during the UC roll-out. However, it is unlikely that the roll-out of UC will have substantial impacts on parental employment or the childcare market for several reasons. First, the main purpose of UC is to replace a collection of ongoing methods of support (albeit with some additional conditionality conditions). Second, the use of the childcare element in the ongoing tax credits is relatively low and there is no reason to believe it will be greater in UC. Third, there is no evidence that previous changes in the proportion of childcare support between 70 percent and 85 percent had any notable impacts.

Second, and most critically, is the concurrent timing of the introduction of TFC and the FEEE 30 hour extension during 2017. Separate impacts can potentially be identified using the fact that the FEEE is only available to a subset of children who are potentially eligible for support through TFC. However, as discussed in the previous subsection, spillover effects of the FEEE to non-eligible ages of children may be important for childcare market outcomes. Consequently, the concurrent introduction of TFC and the FEEE 30 hour extension place substantial limitations for the evaluation of the impacts on the childcare market.

In addition, it should be noted that there could be other changes in childcare policy during the period covering the introduction of TFC and the FEEE 30 hour extension including in the areas of childcare regulation, childcare market facilitation policies and policies affecting the childcare workforce. Indeed, recent policy changes in these areas could have gradual future effects on childcare market outcomes including the enhanced role of Ofsted in 2013, the introduction

¹³ See, for example, Department for Work and Pensions (2014).

of Childminder Agencies in 2013 and the introduction of Early Years Teacher and Early Years Educator qualifications in 2013 and 2014 respectively. Consideration of the potential impacts from these types of policies would be needed in undertaking the proposed evaluations.

4.3 Concurrent other influences on outcomes

There is also a need to consider contextual factors and longer term trends that could affect the outcome measures. For example, these could include:

- The longer term trend of rising employment rates for mothers, including differential trends for mothers in couples and lone mothers.
- Cyclical trends in employment.
- Changes in the availability of informal childcare options (such as grandparents) due to changes in retirement ages and increased working at older ages.
- Perceptions concerning the value of early education for non-work reasons such as child development.
- Rising pressure on space for pre-school childcare in schools as larger cohorts of children reach school age.
- The declining role of Children's Centres in delivering childcare.
- Rising costs of resources needed to deliver childcare including the future implementation of the National Living Wage which could have substantial effects on average pay levels in the childcare sector.

Whether such contextual factors and longer term trends have similar effects on both the policy eligible group and the comparison group should be considered when the evaluation is undertaken.

4.4 Potential size of impacts

In general, it is more challenging to design an evaluation which can robustly identify impacts if the size of impact is potentially small. The evaluations considered here raise several concerns that potential impacts may be small or challenging to identify:

- The subsidy rate in TFC is relatively low at only 20 percent.
- The FEEE is only available for a relatively short period and typically only for one child at a time in a family.
- The FEEE 30 hour extension is an extension of an existing policy and may have smaller impacts than if introduced as a new policy.

Overarching considerations for evaluation approaches

- Potential effects for some metrics are small because participation levels are already relatively high, for example, use of formal childcare among three year and four year old preschool children or employment rates among mothers with only school age children.
- Some childcare market metrics (particularly prices) have exhibited considerable variance over recent years, making the detection of an unusual change more difficult.
- Previous evidence suggests that childcare subsidies tend to have small impacts on parental employment (see section 2.4 above).

4.5 Timing of impacts

The timing of impacts is extremely uncertain. There could be considerable delay in responses given the time required to obtain employment and organise new childcare arrangements, while the FEEE has the added complication that providers must first offer the free places. On the other hand, there is a possibility that childcare providers could anticipate any change in demand resulting from the introduction of TFC and alter their provision or prices prior to changes in demand for childcare taking effect.

4.6 Criteria for assessment

Several criteria were considered in the assessment of the evaluation options:

- What question is being answered?** This considers which impact metrics are evaluated and whether the measure is of the impact of intention to treat (ITT) (i.e. impact on the potentially eligible population including those who do not take-up TFC or the FEEE) or impact on the treated (i.e. impact on those who actually take-up TFC or the FEEE).
- What is the evaluation methodology?** This considers how the impact is being identified and defines the treatment and comparison groups.
- How robust is the approach?** This considers whether conditions for robustness of the methodology have been fulfilled and whether the impacts can be separately identified from other policies or influences.
- What is the best data source to use?** This considers the relative merits and drawbacks of different data sources including the availability of data on the outcomes; the sample sizes and the minimum size of impact that can most likely be identified (the minimum detectable effect (MDE)); the reliability of the data and consistency over time; and the process required to access the data.

- e) **Is sub-group evaluation possible?** This considers whether the analysis can be broken down for single and couple families and by age of children.
- f) **What are the time and monetary costs involved?** This considers the timing of reporting and the costs of obtaining the data and undertaking research.
- g) **What is the overall assessment of the approach?** This considers how likely it is that such an approach will obtain conclusive results which adequately answer the questions addressed.

5 Evaluation of employment impacts

This chapter considers the options for evaluating the impacts on parental employment. It first describes the required impact metrics and available data sources before assessing a number of different evaluation approaches for each policy. It also examines the supplementary roles of using qualitative evidence and a structural modelling approach to simulate impacts. The final subsection makes recommendations on the best approaches.

5.1 Impact metrics

Table 4 presents the impact metrics for an evaluation of the impacts of TFC and FEEE on parental employment.

Table 4. Impact metrics for parental employment

Impact metric	Description
Parental employment	
Mothers' employment rate	Proportion of mothers in work (employment or self-employment).
Mothers' average weekly work hours	Average weekly hours of work for mothers in work.
Mothers' part-time rate	Proportion of mothers working less than 30 hours each week.
Fathers' employment rate	Proportion of fathers in work (employment or self-employment).
Fathers' average weekly work hours	Average weekly hours of work for fathers in work.
Use of formal childcare	
Proportion of children in formal childcare	Proportion of children who attend formal childcare.
Average weekly hours of formal childcare	Average weekly hours of childcare for children attending formal childcare.
Average hourly amount paid for formal childcare	Average hourly amounts paid for children attending formal childcare.

Notes: Mothers and fathers are defined as women and men with natural, adopted or step dependent children living in the same household. A dependent child is one under the age of 16 or aged 16 to 18 and in school or non-advanced further education, not married and living with a parent.

Drawing on the previous evaluations summarised in section 2 and the logic models presented in section 3, the top panel of the table presents the key

employment impact metrics. These are the employment rate and weekly work hours, subdivided into metrics for mothers and fathers. The gender subdivision is required because the potential impacts for mothers is much greater than for fathers due to a lower proportion of mothers currently being in work and previous evidence indicating that childcare policies may have no impact on work behaviour for men. In addition to weekly hours, the metrics for mothers include a metric for the dichotomous division into part-time and full-time work. This measure may be of policy interest as substantial proportions of mothers work part-time and part-time work is sometimes argued to be a barrier to further career progression.

The bottom panel in the table presents the key impact metrics for childcare use including the proportion of children in formal childcare, average weekly hours of childcare and the average hourly amount paid for childcare. These outcomes are not directly within the remit of evaluating impacts on parental employment, but are useful supplementary information in understanding how childcare policy might influence parental work. Hence, the evaluation approaches assessed below consider options which could also evaluate these outcomes.

Three of the potential impacts listed in table 3 in section 2 are not considered in the proposed approaches: impacts on the rate of self-employment; the number of parents looking for work; and reports of unmet demand for childcare. This is because they are secondary, minor potential impacts and of less inherent interest, but could be subject to additional study.

5.2 Data sources

Table 5 summarises the characteristics of the existing UK data sources which could potentially be used for the labour market evaluation.¹⁴ Further details and reviews of other possible data sources are presented in table B1 in Annex B. The potential usefulness of collecting new primary data is considered after the existing data sources have been reviewed.

The best existing data source to evaluate impacts on parental employment is the Labour Force Survey (LFS). This has an extremely large, representative sample and extensive information on employment. Since 1992, it has also contained information on the ages of dependent children, allowing the identification of parents and the analysis of parental employment patterns for over twenty years. The main drawback of this data source is an absence of any information on childcare use which means that no supplementary analysis of the impact on childcare usage can be undertaken using this data source.

¹⁴ A previous review of British data sources on childcare use and maternal employment is presented in Brewer & Shaw (2004).

Table 5. Existing data sources for the employment evaluation

Data source (years)	Sample size per year	Work data	Childcare data	Other data	Caveats
Labour Force Survey (1992+)	300k adults (55k mothers, 45k fathers)	Extensive	None since 2009	Extensive	No childcare data
Family Resources Survey (1994+)	35k adults (6k mothers, 4.5k fathers)	Work status + hours + earnings	Extensive Use of CCTC, ESC	Extensive	No use of FEEE
Understanding Society (2009+)	50k adults (10k mothers, 8k fathers)	Extensive	Type and hours for each child	Extensive	No childcare spending, no childcare policy
Families and Children Study (2002-2008)	7k families (7k mothers, 5k fathers)	Extensive	Extensive	Extensive	Ceased in 2008 No childcare policy
Childcare and Early Years Survey of Parents (2000+)	6.5k families (6.5k mothers, 5k fathers)	Work status + hours	Extensive Use of FEEE / ESC	Limited	Only England Biannual Inconsistencies in data collection No use of CCTC

Notes: CCTC is the childcare element of Working Tax Credit; ESC is Employer Supported Childcare (vouchers). Childcare policy includes FEEE, CCTC and ESC.

There are three data sources which can provide information on both parental employment and childcare outcomes: the Family Resources Survey (FRS), Understanding Society (US) and the Families and Children Study (FACS). All three have similar sample sizes¹⁵ (all considerably smaller than the LFS) and similar childcare information, although US does not currently collect data on childcare spending. The main difference between the surveys is that the FRS is cross-sectional (different households interviewed every year) while the other two surveys are longitudinal (repeated interviews are conducted with the same households) and that the FRS provides a consistent data series since 1994, while US began only in 2009¹⁶ and the FACS only collected data from 2002 to 2008.

¹⁵ Estimates from HMRC suggest that sample size of families in the FRS who would be eligible for TFC is around 1,500 (approximately one quarter of families with dependent children).

¹⁶ US incorporated the sample from the British Household Panel Survey (BHPS) which has a longer data series from 1991, but this subsample is very small as the BHPS was approximately one fifth the size of US.

The main advantage of using either longitudinal surveys is that they permit dynamic analysis (examining changes for the same individuals over time), but has the drawback that panel attrition over time (gradual loss of interviewees) may lead to a less representative sample.

One concern with the FRS is potential under-reporting of the use of formal childcare. This under-reporting was initially documented in Brewer & Shaw (2004) and attributed to wording in the preamble to the questions. Changes to the FRS survey in 2006/7 brought the childcare section into line with the LFS survey, with a key revision being to ask about childcare use in the past seven days rather than in term-time and in school holidays. Current reported levels of formal childcare use are lower in the FRS than in other surveys such as the Childcare and Early Years Survey of Parents.¹⁷ In addition, the most recent FRS data reported only 72 percent of three year olds in formal childcare (see table 13 in Annex C). However, the low proportions in formal childcare reported in the FRS are most likely due to the survey asking about use in the previous seven days rather than during term-time (as used in surveys with higher proportions) and therefore reflect a different measure of childcare use rather than under-reporting.

Finally, it should be noted that while the Childcare and Early Years Survey of Parents provides extensive information on childcare use, it has several significant drawbacks in only collecting data biannually, lacking consistency in the data collection over time and only covering England.

An alternative to using existing data sources would be primary data collection. However, it would be challenging for primary data collection to provide a better alternative data source to evaluate employment impacts because the LFS provides extremely large samples of the required data which is quickly available in the public domain for analysis. But primary data collection could offer alternative sources to the FRS and US for evaluating impacts on childcare use by delivering larger samples with additional information in a more timely fashion. The main drawback of using primary data collection is that it is expensive. The options to use primary data collection are considered within the relevant evaluation methodologies.

5.3 Options for TFC

Three options are considered for the evaluation of the employment and childcare use impacts of TFC:

¹⁷ For example, table 13 in Annex C reports that 32 percent of children under the age of 12 used formal childcare compared to 53 percent of children under the age of 15 (the closest comparable published figure) reported in Huskinson et al (2014).

- A difference-in-difference approach to evaluate the ‘intention to treat’ impacts on parental employment using LFS data and parents with children only over the age of 11 as a comparison group.
- Use of a staggered roll-out across different ages of children to evaluate the ‘intention to treat’ impacts on parental employment using LFS data or on childcare usage using FRS data.
- Use of a matching approach to evaluate the ‘treated’ impacts on parental employment and childcare usage using either US data or primary data collection and parents not taking up TFC as a comparison group.

5.3.1 Difference-in-difference approach using parents with older children

a) *What question is being answered?*

The question being addressed is the ‘intention to treat’ impacts on parental employment. This asks how much TFC raises the employment rate and work hour impact metrics among all parents with children aged 11 or younger, regardless of whether they use TFC or whether they meet the work and childcare requirements to claim TFC. In other words, it considers the impact on parental employment for all parents who are either already eligible for TFC or can change their work and/or childcare choices to become eligible and claim TFC. Hence, it considers the group who are ‘intended to be treated’ but do not necessarily receive the treatment.

b) *What is the evaluation methodology?*

The evaluation methodology is a difference-in-difference approach using parents with dependent children who are only aged over 11 as the comparison group.¹⁸ Parents in this comparison group lie outside the ‘intention to treat’ group because they cannot alter their work and childcare choices to become eligible for TFC, but they are potentially similar in their employment trends as they have older children.¹⁹ The impact of TFC is measured as the difference in changes in employment rates between the treated group and the comparison group over the period that TFC is introduced between 2016 and 2018.

¹⁸ A dependent child is a natural, adopted or step child living in the same household who is aged under 16 or aged 16 to 18 and in full-time education.

¹⁹ It is important to note that the treatment and comparison groups are not required to have the same levels in the employment metrics (which does not hold as employment rates are higher among mothers with only older children), but only to have similar or a steady difference in trends in the absence of the change in policy for the treatment group.

c) How robust is the approach?

There are three issues which need to be addressed to ensure the robustness of this approach.

First, it is important to either check that trends in the employment metrics between the two groups would be the same in the absence of the introduction of TFC or to control for ongoing differences in trends. For example, if employment rates among mothers with children aged 11 or under were rising faster than those for mothers with older children in the years preceding the introduction of TFC, this difference might be expected to persist during the period of the introduction of TFC and should be subtracted from the observed difference to obtain the impact of TFC. This is a triple difference or DiDiD (difference-in-difference-in-difference) approach described in section 2. A potential risk to the robustness of the approach is that differences in the trends prior to TFC might have too much variation to precisely estimate what the ‘normal’ difference would have been in the absence of the introduction of TFC. The longer the period over which the ‘normal’ difference can be estimated, the more likely a reasonably precise estimate can be obtained. An examination of the longer trends could be undertaken prior to the introduction of TFC in preliminary evaluation analysis.

Second, in order to test whether any estimated impact is also capturing the concurrent introduction of the FEEE 30 hour extension, it would be necessary to estimate the impact excluding (or with controls for) parents of three and four year old children who could also be affected by the introduction of the FEEE 30 hour extension. This would reduce the sample sizes and have slightly higher minimum detectable effects than those presented below.

Finally, it would be important to test whether 2018 allowed sufficient time for the impacts on employment to occur. Evidence on take-up of TFC could provide some indication on the timing of impacts: while claimant numbers are rising, it is unlikely that the full impacts on employment have occurred. However, the most robust approach would be to repeat any initial analysis using data for 2019 (and possibly subsequent years if TFC take-up continues to rise) to allow for potential delays in impacts.

d) What is the best data source to use?

The LFS is the best existing data source to use for this approach as it provides the required employment metrics for large, representative samples of both the treatment and control groups over the required period. There is also a long series of historic, consistent data over which the longer term trends in the employment metrics can be used to undertake the DiDiD tests.

Power tests estimating the minimum detectable effect indicate that the LFS has a sufficient sample size to detect an impact of approximately 1.1 percentage points for employment rates for both mothers and fathers and of approximately 0.4

Evaluation of employment impacts

hours for average weekly hours of work for both mothers and fathers (see Annex C). Hence, it is likely that this approach using the LFS data would detect impacts which are larger than these thresholds.

e) Is sub-group evaluation possible?

The LFS data collects sufficient information for the evaluation to be undertaken separately for parents in couples and for lone mothers and by different ages of youngest child (including separately for parents with pre-school children and for parents with only school age children). The main drawback is that the thresholds for the size of impacts that can be detected with the smaller subsamples will be higher. In particular, sub-group evaluation for lone mothers may only detect much larger impacts given the much smaller number of lone mothers than mothers in couples. Sub-group analysis by age of youngest child might also provide more precise estimates in the longer term ‘normal’ patterns in the employment metrics. For example, parents with only school-age children in the treatment groups might have a closer match in trends with parents with children only over the age of 11.

f) What are the time and monetary costs involved?

Based on current plans for the introduction and roll-out of the policy, analysis of the LFS data should be possible in 2019²⁰ and initial findings potentially available within that year. In the absence of initial conclusive findings of a positive impact, it would be necessary to repeat the analysis in at least one further year with findings in 2020 in order to test for delayed impacts.

LFS data is publicly available at no cost for government-funded projects, so the only monetary costs would be for the analysis. An evaluation of this nature might cost in the region of £70k to £120k to undertake, although precise costs would depend upon the specification of the reporting requirements and the nature of the organisation undertaking the work.²¹

g) What is the overall assessment of the approach?

Overall, this approach is likely to identify impacts of a reasonably small size in a robust manner. It can differentiate impacts from those of the concurrent introduction of the FEEE 30 hour extension, assuming that the FEEE 30 hour extension has negligible spillover effects to parents who do not have children eligible for that policy. The main drawback is that it cannot provide any additional insight into the impacts on childcare usage.

²⁰ LFS data for 2018 is likely to be available by the end of the first quarter of 2019.

²¹ Throughout, costs are approximate estimates based on typical guide amounts presented in Government “Initiations to Tender” for this type of work. The key distinction in cost is between approaches requiring secondary analysis of existing data sources and those requiring the collection of primary data.

5.3.2 Staggered roll-out across ages of children

a) *What question is being answered?*

The question being addressed is the ‘intention to treat’ impacts on parental employment and childcare use. This asks how much TFC raises employment rate and work hour impact metrics and how much TFC raises childcare use for each age of child under the age of 12 (again, regardless of whether they use TFC or whether they meet the work or childcare requirements to claim TFC).

b) *What is the evaluation methodology?*

The approach would use a comparison of the change in parental employment and childcare use between parents with the youngest child of an age just becoming eligible for TFC and parents with other children aged under 12 during the roll-out period.

c) *How robust is the approach?*

The period of roll-out would need to be sufficiently long to allow sufficient time for impacts to have taken effect between stages and to ensure that any anticipation effects did not mask any impacts. For example, roll-out on a monthly or bimonthly basis would be unlikely to be sufficient time to ensure this. Longer periods in the staggering of the introduction would be required to achieve robust findings, but the minimum required length depends upon the unknown time required for impacts to take effect. The potential benefits of blocking the roll-out into broader ages of children also depends upon unknown differences in impacts across ages. An approximate estimate of a robust stagger period might be a minimum of six months to allow families time to find new work opportunities and childcare arrangements, although a more robust length would be one year given the evidence on the length of response in mothers’ employment to children starting school at age four. If a six month (one year) roll-out was staggered into five blocks of ages (for example, two pre-school and three school age groups), this would require a roll-out period of two years (four years) and evaluation period of at least two and a half years (five years).

As with the previous approach, allowance would need to be made for any ongoing differences in the trends between the different ages of youngest child (a DiDiD approach). This could be estimated using differences in months without TFC being rolled out to the age groups within the roll-out period

d) *What is the best data source to use?*

The LFS would provide the best data source to evaluate the employment impacts, but the treatment samples for each age group would be small (for example, a monthly roll-out would have samples of less than 500) and only larger impacts could be identified. The FRS would provide the best data source to

Evaluation of employment impacts

evaluate childcare use impacts, but the treatment samples for each age group would be extremely small. Collection of primary data could generate larger sample sizes, but would require a large number of surveys over a short period.

e) Is sub-group evaluation possible?

By construction, this approach would consider the impacts for different ages of children. However, disaggregation for couples and lone mothers would have extremely small sample sizes and would be unlikely to generate robust findings.

f) What are the time and monetary costs involved?

If roll-out of the policy by age of youngest child were undertaken on a six-month stagger, analysis of the employment impacts for the age groups in the initial stages of roll-out using LFS data should be possible in early 2018 and findings potentially available within that year. Analysis of the childcare use impacts using the FRS should be possible in 2019 with findings available within that year.²²

Both LFS and FRS data are publicly available at no cost for government-funded projects, so the only monetary costs would be for the analysis. An evaluation of this nature might cost in the region of £100k to £130k to undertake, although precise costs would depend upon the specification of the reporting requirements and the nature of the organisation undertaking the work. Primary data collection would have substantially higher costs, particularly in order to undertake an extremely large number of surveys within a short time period.

g) What is the overall assessment of the approach?

Unless any roll-out takes place over a suitably long period to ensure there is sufficient time for impacts to take effect and to permit sufficiently large sample sizes to be drawn from the data, this would not be a robust evaluation approach.

5.3.3 Matched panel based on TFC take-up

a) What question is being answered?

The question being addressed is the actual treatment effect of taking up TFC on parental employment and childcare use. This asks how much TFC raises the employment rate and work hour impact metrics and how much TFC raises childcare use for parents who take up TFC in comparison to otherwise similar parents who do take-up TFC.

²² FRS data is collected on an annual basis between April and March but the data is typically not publically available until at least a year later. Hence, data for 2017 will be available in 2019.

b) What is the evaluation methodology?

The evaluation methodology uses a comparison of changes in employment and childcare metrics over the period of TFC introduction (2016 to 2018 (or possibly 2019)) between parents who take up TFC and potentially eligible parents who do not take up TFC. The comparison group would be identified from all parents with a child aged 11 or younger not taking up TFC either because they do not fulfil the work or childcare use requirements or because they fulfil the requirements but choose not to claim TFC. A matching technique would be used to ensure that the comparison group resembles those taking up TFC as closely as possible in characteristics related to employment and childcare use other than the receipt of TFC. For example, the matching technique could be the use of control variables in regression analysis or propensity score matching to extract a specific comparison group based upon factors related to the outcome metrics at the baseline in 2016²³.

c) How robust is the approach?

The robustness of this approach is reliant upon the existence of some factor or set of factors which influence TFC take-up but are unrelated to the employment or childcare use impact metrics. Take-up of policies of this type is rarely complete for a mixture of reasons that can be unrelated to work or childcare choices. For example, take-up of TFC could be incomplete because providers or parents do not consider the subsidy amount worth the time and effort of claiming. Or parents may not claim because of social stigma; a general dislike or distrust of government sources of support; or a lack of awareness of the policy (possibly related to local levels of advertising). The number of comparable individuals not taking up TFC would also need to be of sufficient size for robust comparisons to be drawn. A priori, it is not possible to estimate whether or why take-up may not be complete which means that this evaluation strategy carries a high risk of failing to be able to identify impacts.

d) What is the best data source to use?

The best existing data source to use would be US which contains the required panel element of repeated interviews with the same individuals over time. However, an additional question on the take-up of TFC would need to be added to the survey from 2017 and also questions on childcare spending from 2016 if impacts on hourly childcare expenditure are to be estimated.

²³ For example, the control variables for the employment models could include age of youngest child and number of children; parents' age, highest qualification, ethnicity, housing type and longstanding health problem or disability; and measures of local labour market conditions. The control variables for the childcare outcomes could include similar factors plus the availability of informal childcare options and urbanity and region to capture local childcare availability.

Assuming a 50 percent TFC take up rate, power tests estimating the minimum detectable effect indicate that the US has a sufficient sample size to detect minimum impact sizes of approximately 4.8 percentage points for the employment rate for mothers and 4.2 percentage points for fathers and of approximately 1.5 hours for average weekly hours of work for both mothers and fathers (see Annex C). Under the same assumption, the minimum detectable effect for the proportion of children using formal childcare is estimated to be 3.6 percentage points. A higher take-up rate reduces the size of the comparison group and makes it more difficult to detect smaller impacts. For example, it is estimated that a take-up rate of 90 percent increases the estimated minimum detectable impacts for employment rates to 9.4 and 6.2 percentage points for mothers and fathers respectively and for weekly work hours to 2.3 and 2.5 hours respectively. It raises the estimated minimum detectable impact to 6.0 percentage points for the proportion of children in childcare. Although these impacts are large relative to existing evidence on the potential sizes of impacts, it should be considered that these are impacts for those actually treated (i.e. taking up TFC) rather than the entire potentially eligible group (i.e. intention to treat) typically considered in existing evidence. Hence, it is more difficult to gauge whether actual impacts are likely to exceed these minimum sizes of detectable effects.

An alternative data source could be the use of primary data collection using a bespoke panel of three survey waves with parents of children under age 12, ideally undertaken just prior to the introduction of TFC and around 6 months and 18 months after the end of the roll-out. Such a survey could collect data on the parental employment and childcare use outcome metrics, but also on the use of different childcare policies and reasons for take-up and non-take-up. The latter would be potentially useful in identifying whether the driver of take-up was independent of the impact metrics. In addition, the survey sample could be weighted towards families more likely to use TFC, that is, actual or potential higher earners using an initial screening of potential interviewees.

Assuming a bespoke survey of three waves with 6,000 households (double the size of the US data) and a potential equal division of respondents into those taking up TFC and a comparison group, power tests estimating the minimum detectable effect indicate that such a bespoke survey would be able to detect an impact of approximately 3.4 percentage points for employment rates for mothers and 2.5 percentage points for fathers and of approximately 1.1 hours for average weekly hours of work for mothers and 0.9 hours for fathers (see Annex C). Under the same assumption, the minimum detectable effect for the proportion of children using formal childcare is estimated to be 2.5 percentage points. A division of the sample into 90 percent taking-up TFC and 10 percent not taking up would, as before, increase the size of the estimated minimum detectable effects, although to a level around the size of or just greater than those for the US sample with 50 percent take-up (see table 15 in Annex C). With bespoke data collection, it might also be possible to use screening at the second stage to

enhance the proportion of those not taking-up TFC and hence reduce the minimum sizes of impacts that can be detected.

e) Is sub-group evaluation possible?

Given the modelling demands of matching the sample using observed characteristics, it is unlikely that sub-group evaluations would be possible with the US data. Primary data collection might be designed with booster subsamples (more respondents) for subgroups of particular interest, for example, it could focus only on pre-school children.

f) What are the time and monetary costs involved?

Based on current plans for the introduction and roll-out of the policy, analysis of the US data should be possible in 2019/2020²⁴ and initial findings potentially available within 2020. As with the first approach considered, in the absence of initial conclusive findings of a positive impact, it would be necessary to repeat the analysis in at least one further year with findings in 2021 in order to test for delayed impacts, particularly if initial claimants of TFC were those who did not need to alter their work or childcare choices to be eligible.

US data is publicly available at no cost for government-funded projects, so the only monetary costs would be for the analysis. An evaluation of this nature might cost in the region of £70k to £120k to undertake, although precise costs would depend upon the specification of the reporting requirements and the nature of the organisation undertaking the work.

The alternative data source using primary data collection could potentially generate findings more quickly, in part because a smaller, more focused survey might need less time to be processed. But the monetary costs would be considerably higher: a bespoke survey of three waves with around 6,000 parents could cost in the region of £1.5 to £3 million.

g) What is the overall assessment of the approach?

This approach is risky as it relies on sufficient variation in the take-up of TFC among otherwise similar groups of parents to identify impact and this means that feasibility cannot be assessed prior to policy introduction. In addition, any findings could be subject to the criticism that the reasons driving variation in take-up have not been adequately allowed for unless there is evidence of a driver of take-up which is not related to the impact metrics. Use of primary data collection could potentially identify smaller impacts than use of the existing US

²⁴ US data is collected on a rolling wave basis, each wave covering just over two years with data publicly available around 6 months after the end of the fieldwork period. Hence, half of the data for 2018 will be available in wave 9 in late 2019 and half in wave 10 in late 2020.

data, but only at substantial cost for a relatively small gain in the minimum detectable sizes of impacts.

5.4 Options for the FEEE 30 hour extension

Three options are considered for the evaluation of the employment and childcare use impacts of the FEEE 30 hour extension. All three options evaluate the ‘intention to treat’ impacts on parental employment (using LFS data) and on childcare use (using FRS data). The three methodologies are:

- A difference-in-difference approach using parents with younger children (aged 1-2) and older children (aged 5-7) as comparison groups.
- A difference-in-difference approach using any variation in the speed of implementation in the offer of free places.
- A random discontinuities design exploiting the cut-off for eligibility at age three in comparison with earlier periods when the FEEE 30 hour extension was not available.

5.4.1 Difference-in-difference with parents with younger and older children

a) *What question is being answered?*

The question being addressed is the ‘intention to treat’ impacts on parental employment and childcare use. This asks how much the FEEE 30 hour extension raises the employment rate and work hours among all parents with three and four year old per-school children and alters their childcare choices, regardless of whether they use the FEEE 30 hour extension or whether they meet the work requirements to be eligible FEEE 30 hour extension.

b) *What is the evaluation methodology?*

The evaluation methodology is a difference-in-difference approach using parents with younger children (aged 1-2) and no child aged three and parents with older children (aged 5-7) and no child aged three. Parents in these comparison groups lie outside the ‘intention to treat’ because their children are not eligible for the FEEE 30 hour extension, but they are potentially similar in their employment and childcare use trends.²⁵ The impact of the FEEE 30 hour extension is measured as the difference in changes in the employment and childcare use

²⁵ It is important to note that the treatment and comparison groups are not required to have the same levels in the employment and childcare impact metrics (which does not hold as employment rates are higher among mothers with only older children), but only to have similar or a steady difference in trends in the absence of the change in policy for the treatment group.

metrics between the treated group and the comparison groups over the period that the FEEE 30 hour extension is introduced between 2016 and 2018.

c) How robust is the approach?

This approach is analogous to the first option considered for the evaluation of employment impacts for TFC and faces three similar issues in ensuring robustness.

First, it is important to either check that trends in the employment and childcare use metrics between the treatment and each comparison group would be the same in the absence of the introduction of FEEE 30 hour extension or to control for ongoing differences in trends. Any ongoing differences in trends should be subtracted from the observed difference during the FEEE 30 hour extension period to obtain the impact of the extension. This is essentially a triple difference or DiDiD (difference-in-difference-in-difference) approach. The longer the period over which the ‘normal’ difference in ongoing trends can be estimated, the more likely a reasonably precise estimate can be obtained. An examination of the longer trends could be undertaken prior to the introduction of the FEEE 30 hour extension in preliminary evaluation analysis.

Second, in order to simply identify the impact of the FEEE 30 hour extension separately from the concurrent introduction of TFC, TFC would need to have similar impacts for the treatment and each comparison group. However, it is possible that the impacts of TFC could differ by age of child and the impacts for three and four year old preschool children could differ from those for children in the two comparison groups. Using both a younger and older comparison group allows an assessment of the extent to which differential impacts of TFC by age could affect the estimated impacts for the FEEE 30 hour extension.²⁶

Finally, it would be important to test whether 2018 allowed sufficient time for the impacts on employment and childcare use to occur, particularly as there is an initial step where childcare providers must first offer the FEEE 30 hour places. Evidence on take-up of the FEEE 30 hour places could provide some indication on the timing of impacts: while the proportion of children using such places is rising, it is unlikely that the full impacts on employment would have occurred. The most robust approach would be to repeat any initial analysis using data for 2019 (and possibly subsequent years if the use of the FEEE 30 hour places continues to rise) to allow for potential delays in impacts.

²⁶ For example, the impact of TFC could be considered to change in a simple linear manner with age of child and the average impact for the younger and older comparison groups could be subtracted from the total difference for three and four year old preschool children to obtain an estimate of the impact of the FEEE 30 hour extension. Or the FEEE 30 hour extension eligible age group could be compared only with the younger age group under the scenario that TFC impacts all pre-school children in a similar way or only with the older age group under the scenario that TFC impacts three and four year old preschool children in a similar way to younger children at school.

In addition, it is important to note that this approach assumes that spillover effects of the FEEE to parents without three or four year old preschool children are negligible.

d) What is the best data source to use?

The LFS is the best data source to use for the evaluation of employment impacts as it provides the required metrics for reasonably sized, representative samples of both the treatment and control groups over the required period. There is also a long series of historic, consistent data over which the longer term trends in the employment metrics can be used to undertake the DiDiD tests. Power tests estimating the minimum detectable effects indicate that the LFS has a sufficient sample size to detect an impact of approximately 2.5 percentage points for the employment rate for mothers and 1.7 percentage points for fathers and of approximately 0.8 hours for average weekly hours of work for mothers and 0.6 hours for fathers (using the comparison with younger children) (see Annex C).

The FRS is the preferable data source over US to estimate impacts on childcare use because it has the advantages of containing data on childcare spending; of having a long historic series of data to estimate longer term trends and of having no risk of any sample bias from panel attrition. Power tests estimating the minimum detectable effect indicate that the FRS has a sufficient sample size to detect an impact of approximately 7.7 percentage points for the proportion of children participating in formal childcare (again, using the comparison with younger children). These are quite large impacts, but the key impact of interest for the FEEE 30 hour extension could be the weekly hours of childcare rather than participation.²⁷

e) Is sub-group evaluation possible?

As the FEEE policy is targeted on three and four year old preschool children, sub-group evaluation is only of potential interest for parents in couples and lone mothers, but is unlikely to be feasible due to insufficient sample sizes.

f) What are the time and monetary costs involved?

Based on current plans for the introduction and roll-out of the policy, analysis of the LFS and FRS data should be possible in 2019 and 2020 respectively²⁸ and initial findings potentially available within those years. In the absence of initial conclusive findings of a positive impact, it would be necessary to repeat the

²⁷ The data required to estimate the minimum detectable impacts for weekly hours of childcare involves considerably greater preparation than that for participation in formal childcare and is not possible within the scope of this report.

²⁸ LFS data for 2018 should be available in early 2019, while FRS data for 2018 will not be available until 2020.

analysis in at least one further year with findings in 2020 and 2021 in order to test for delayed impacts.

Both LFS and FRS data are publicly available at no cost for government-funded projects, so the only monetary costs would be for the analysis. An evaluation of this nature might cost in the region of £100k to £130k to undertake, although precise costs would depend upon the specification of the reporting requirements and the nature of the organisation undertaking the work.

g) What is the overall assessment of the approach?

Overall, this approach is likely to identify employment and childcare use impacts of a reasonable size in a robust manner. There are two main caveats. First, separately identifying impacts from those of the concurrent introduction of TFC may require consideration of a range of estimates if TFC has differential impacts by age of child. Second, the approach assumes that spillover effects of the FEEE to parents without three or four year old preschool children are negligible.

5.4.2 Difference-in-difference using variation in the speed of implementation

a) What question is being answered?

As with the previous methodology, the question being addressed is the ‘intention to treat’ impacts on parental employment and childcare use. This asks how much the FEEE 30 hour extension raises the employment rate and work hours among all parents with three and four year old preschool children and alters their childcare choices, regardless of whether they use the FEEE 30 hour extension or whether they meet the work requirements to be eligible for the FEEE 30 hour extension.

b) What is the evaluation methodology?

The evaluation methodology is a difference-in-difference approach between geographic areas where FEEE 30 hour places become available sooner and areas where the policy is slower to be implemented. The impact is measured as the difference in change in employment and childcare use metrics for parents with three and four year old preschool children where the policy has been implemented with those for similar parents where the policy has not been implemented.

c) How robust is the approach?

There are two potential problems with this approach. First, it is reliant on the existence of geographical variation in the implementation of the FEEE 30 hour places and it is unknown whether this will be the case prior to the policy introduction. Second, the estimated impacts will be biased if the reason for

Evaluation of employment impacts

variation in the speed of implementation is related to geographical differences in employment and childcare use.

d) What is the best data source to use?

For similar reasons as in the previous approach, the LFS and FRS are the best data sources for this approach. However, the sample is limited to parents of three and four year old preschool children, divided into two smaller subsamples of treated and comparison areas and the minimum detectable impacts would be correspondingly large. Data would also be required on geographical variation in the implementation of the offer of the FEEE 30 hour places.

e) Is sub-group evaluation possible?

As the FEEE policy is targeted on three and four year old preschool children, sub-group evaluation is only of potential interest for parents in couples and lone mothers, but is unlikely to be feasible due to insufficient sample sizes.

f) What are the time and monetary costs involved?

The time required for the evaluation would be dependent on how long it took for a sufficient proportion of areas to have implemented the policy to form a sufficiently large treatment group. In the ideal scenario, analysis of the LFS and FRS data would be possible in 2019 and 2020 respectively and initial findings potentially available within those years. Again, in the absence of initial conclusive findings of a positive impact, it would be necessary to repeat the analysis in at least one further year with findings in 2021 in order to test for delayed impacts.

Both LFS and FRS data are publicly available at no cost for government-funded projects, so the only monetary costs would be for the analysis. An evaluation of this nature might cost in the region of £100k to £130k to undertake, although precise costs would depend upon the specification of the reporting requirements and the nature of the organisation undertaking the work.

g) What is the overall assessment of the approach?

This approach contains a high risk of failing to generate conclusive findings for two reasons. First, it is dependent upon there being substantial variation in the speed of implementation. Second, the limited sample sizes mean only larger impacts would be identified.

5.4.3 Random discontinuities design exploiting the age cutoff for eligibility

a) What question is being answered?

As with the previous two methodologies, the question being addressed is the 'intention to treat' impacts on parental employment and childcare use. This asks how much the FEEE 30 hour extension raises the employment rate and work

hours among all parents with three and four year old preschool children and alters their childcare choices, regardless of whether they use the FEEE 30 hour extension or whether they meet the work requirements to be eligible for the FEEE 30 hour extension.

b) What is the evaluation methodology?

The evaluation methodology is a random discontinuities design exploiting the cut-off for eligibility at age three in comparison with earlier periods when the FEEE 30 hour extension was not available. The impact of the FEEE 30 hour extension is estimated by comparing the employment and childcare metrics for parents just prior to the age at which their child becomes potentially eligible for an FEEE 30 hour place with those for parents shortly after their child becomes potentially eligible. The comparison over the short timeframe presumes that conditions for the two sets of parents are identical except for the change in FEEE eligibility. The separate impacts of FEEE 30 hour extension from the impacts of the ongoing 15 hour FEEE offer are estimated as the difference in the comparisons between the periods prior to and after the introduction of the FEEE 30 hour extension.²⁹

c) How robust is the approach?

This approach is reasonably robust. The main caveat is the need to assume that there are no anticipation effects of becoming eligible for FEEE and that there is relatively little delay in impacts. In addition, the approach requires that there is no other reason for a sudden change in employment and childcare choices at the same point in time and this appears to hold. Finally, the concurrent introduction of TFC is not an issue as it is reasonable to assume that it will have very similar effects for families either side of the FEEE eligibility cut-off point.

d) What is the best data source to use?

Again, the best data sources to use are the LFS for employment impacts and the FRS for impacts on childcare use. Information on children's month of birth would be required which is available using secure access for the LFS^{30,31}. While sample sizes of parents close to the eligibility cut-off points are small, the approach can use several years of data prior to and post the extension as it does

²⁹ This approach is a type of "difference-in-difference" approach which the first difference estimated using the random discontinuities design.

³⁰ Prospective users of a Secure Access version of the LFS need to fulfil additional requirements, including the completion of an extra application form to demonstrate why they need access to the extra, more detailed variables; face-to-face training; and agreement to a Secure Access' User Agreement and Breaches Penalties Policy.

³¹ Month of birth for children is also likely to be available for the FRS under similar conditions.

Evaluation of employment impacts

not use information on changes in employment and childcare use over the period of introduction. Combining three years of data would generate estimated minimum detectable effects of a reasonable size. An alternative approach to obtain larger sample sizes in the FEEE 30 hour extension period would be to temporarily boost the subsamples of parents with children of the relevant age in the LFS and FRS in 2018, possibly using a shortened survey for the booster sample focused on collecting data for this evaluation. The booster would likely need to treble the size of the relevant subsample to obtain estimated minimum detectable effects of a reasonable size.

Power tests estimating the minimum detectable effects indicate that one year of the LFS has a sufficient sample size to detect an impact of approximately 3.9 percentage points for the employment rate for mothers and 2.6 percentage points for fathers and of approximately 1.1 hours for average weekly hours of work for mothers and 0.9 hours for fathers (see Annex C). A longer period of LFS data (three years) reduces these estimated minimum detectable impacts to 2.2 and 1.6 percentage points for the employment rates for mothers and fathers respectively and to 0.6 and 0.5 weekly work hours respectively. As with the initial methodology considered above, the minimum detectable effects for the proportion of three year old children in formal childcare using FRS data are quite high: 10.5 percentage points using one year of data and 6.3 percentage points using three years of data. Again, however, the primary impact of interest for FEEE 30 hour extension may be the weekly hours of childcare rather than participation.

e) Is sub-group evaluation possible?

As the FEEE policy is targeted on three year old and four year old preschool children, sub-group evaluation is only of potential interest for parents in couples and lone mothers, but is unlikely to be feasible due to insufficient sample sizes.

f) What are the time and monetary costs involved?

As several years of data are required after the introduction of the FEEE 30 hour extension, findings would not be available until at least 2021, based on current plans for the introduction and roll-out of the policy. Both LFS and FRS data are publicly available at no cost for government-funded projects, so the only monetary costs if there were no boosts to the samples would be for the analysis. An evaluation of this nature might cost in the region of £100k to £130k to undertake, although precise costs would depend upon the specification of the reporting requirements and the nature of the organisation undertaking the work.

If booster samples were introduced into the surveys, findings could be available earlier, potentially in 2019. There would be an additional monetary cost for the booster samples, but an estimate of the size of this cost would need to be obtained from those currently undertaking the survey.

g) What is the overall assessment of the approach?

Overall, this approach is likely to identify employment and childcare use impacts of a reasonable size in a robust manner. The main caveat is the concern that impacts need to occur with relatively little delay. In addition, findings would only be available with a considerable time lag unless booster samples were added to the LFS and/or FRS surveys but this could involve considerable cost.

5.5 Options for the FEEE previous extensions

As noted in section 2, an evaluation of the impacts of the extension of the FEEE to three year olds from 2004 on maternal employment has been undertaken using LFS data (Brewer et al (2014)). This evaluation used two methodological approaches. First, a difference-in-differences approach using variation in the speed of implementation across geographic areas and across ages of children (a combination of the approaches proposed in subsections 5.4.1 and 5.4.2). Second, a regression discontinuities design using eligibility cut-offs with mother fixed effects (akin to the approach proposed in subsection 5.4.3). Ongoing work by Brewer et al is refining the employment impact analysis for the second approach and is also extending the evaluation to include an evaluation of the impacts on childcare use using FRS data.

Similar approaches using this data could also be used to evaluate the employment and childcare use impacts of the extensions of the FEEE to the most disadvantaged families in 2013 and 2014. For the first approach based on geographical variation in the speed of implementation, although there has been geographic variation in the take-up of the two year-old offer, it is not clear whether this is related to substantial variation in the implementation of making free places available. In addition, this approach could also run into problems by overlapping with further related policy changes in 2017. A serious limitation to using the random discontinuities design is the much smaller number of families who are eligible for the two year-old offer in comparison to the offer for three and four year olds which would mean that a larger number of years of data would be required to generate sufficient sample sizes to identify impacts of smaller sizes.

An alternative option would be to consider using data collected as part of the Study of Early Education and Development (SEED) on the employment and childcare choices for parents of two to four year olds over the period 2013 to 2016. Exploratory analysis will be undertaken by the team conducting the SEED evaluation on whether the data can be used to evaluate the impacts of the FEEE policy for two year olds on parental employment and childcare use.

Overall, the best evaluation option for the previous extensions to the FEEE are already either being undertaken or explored.

Evaluation of employment impacts

5.6 Role of qualitative evidence

Collection of qualitative evidence on the impacts of TFC or the FEEE 30 hour extension could potentially add two types of information. First, it could provide additional evidence on the impacts on parental employment and childcare use considered in the quantitative evaluation approaches above. Second, it could compliment the findings of the quantitative approaches by providing information to help understanding of the reasons and mechanisms for the findings from the quantitative approaches.

To supplement the findings from the quantitative evaluation, a qualitative approach might ask parents how they thought the introduction of TFC or the FEEE 30 hour extension had affected their work and childcare choices. However, there are several caveats on the value of these responses. First, perceptions of choices in the hypothetical counterfactual may not match actual choices in the alternative situation or may only be probabilistic responses which are hard to quantify. Second, assessing changes in choices over the period of policy change is complicated by the ongoing changing context brought about by the natural aging of children. Third, parents of three and four year olds might find it challenging to separately identify the impacts of TFC and the FEEE 30 hour extension.

The second approach could gather evidence on the reasons for parents' work and childcare choices using either individual qualitative interviews or focus groups of parents. This could investigate the reasons for decisions concerning take-up of both policies (including such process issues as awareness and application cost) and why the decision to take-up had or had not involved changes in work or childcare choices. It could also examine the relative importance of some of the other influences identified in the logic models in section 2 including the role of employment opportunities, availability of childcare and parental preferences between formal and informal types of childcare. Because qualitative analysis of this type of data would provide explanations for some types of behaviour rather than quantifying the extent of behavioural impacts, large samples would not be required and the cost would be accordingly moderate. It would be important, however, to ensure that the data collection was not undertaken too early before the policies have had time to 'bed down' and at least initial impacts felt.

Overall, qualitative approaches would be most useful to provide complimentary, explanatory information for the findings from the quantitative approaches rather than as supplementary evidence on the impacts. In particular, they are useful for enhancing understanding how policy design might be improved and the limitations of the impacts of the policy due to other influences on work and childcare choices.

5.7 Supplementary structural modelling

A second complimentary analysis to the quantitative evaluations could be the use of structural modelling to estimate the potential employment and childcare use responses to TFC and the FEEE 30 hour extension.

Structural modelling essentially simulates policy impact from current information about individuals' choices rather than observing the impacts of a policy. In this case, current data on parental employment and childcare use from the FRS³² would be used to estimate how parents respond to variation in childcare prices and these estimated responses used to predict the impacts of the childcare subsidy in TFC and the free places in the FEEE 30 hour extension. Sample sizes could be enhanced by combining data from more years of the FRS to help produce more precise estimates.

This modelling could be undertaken immediately and provide findings prior to the implementation of policy. It would also be a relatively low cost option with costs only for analysis, potentially in the region of £70k to £120k, although precise costs would depend upon the specification of the reporting requirements and the nature of the organisation undertaking the work.

However, such structural modelling and simulation of the impacts of childcare policy must address several technical issues:

- The modelling must identify a source of 'genuine' price variation which takes into account differences in the quality of childcare. The ideal is quality-adjusted prices from providers, but this is not available and difficult to obtain. Typically, the literature has measured price variation as average hourly spending reported by families within geographic localities or predicted prices based on supply factors such as local regulations and workforce, but only a few studies have included quality-adjusted measures of hourly cost.³³
- The absence of wage rates for non-workers and childcare prices for those not using childcare is usually addressed using estimates from selection-corrected wage and price regressions.
- Allowance should be made for the availability of informal, unpaid childcare options, but this is typically not addressed due to a lack of data on those options.

³² The larger LFS would not be suitable because the data source must contain some information on childcare prices or hourly amounts spent on childcare to proxy price.

³³ For further discussion and an example of modelling using quality-adjusted measure, see Paull, Taylor & Duncan (2002).

In addition, simulations for TFC and the FEEE 30 hour extension would require particularly sophisticated modelling because neither policy is a simple, flat-rate subsidy for childcare costs. TFC is complicated by the implicit lower earnings threshold (and alternative subsidy in the WTC/UC), while the FEEE policies provide a discontinuous lump-sum amount of free childcare (within a limited timeframe).

Most importantly, simulation of policy responses may overstate impacts for several reasons:

- Responses to a childcare subsidy may differ from those to price variation due to incomplete take-up or limited or uncertain duration of the subsidy.
- The simulations make no allowance for constraints on employment opportunities or childcare availability which could inhibit parents from realising their preferred options.
- The simulations make no allowance for potential second round reactions from childcare providers who may increase prices in response to greater demand.

Overall, such modelling can usefully provide an indication of the potential ‘upper bound’ of the size of impacts, but it is no substitute for an evaluation of actual impacts.

5.8 Recommendations

Several different options to evaluate the impacts of TFC and FEEE on parental employment and childcare use have been assessed. This subsection highlights which approaches, on balance, could be considered the best approaches to be used.

The recommendations for the evaluation of TFC are:

- To evaluate the ‘intention to treat’ impacts on parental employment using a difference-in-difference methodology with a comparison group of parents of dependent children aged over 11 applied to LFS data. This approach has a high probability of identifying conclusive findings, can be undertaken at reasonably low cost and, based on current plans for the introduction and roll-out of the policy, could produce findings by 2019 (or 2020 if impacts take longer to materialise).
- To (optionally) consider the evaluation of the ‘treatment’ impacts on childcare use using a matched panel methodology based on TFC take-up applied to US data. Although this approach carries a substantial risk that variation in TFC take-up may be insufficient to identify conclusive

findings, the relatively low cost of the analysis of US data mean that the approach is worth considering.

The feasibility of evaluating impacts of TFC on parental employment is driven by two factors. First, the existence of a comparable group of parents who face a similar employment context but are not eligible for TFC because their children are too old. Second, the availability of a large and robust data source with the required employment metrics. Evaluating impacts on childcare is more challenging because, while parents with only older children (over age 11) are a reasonable comparison group for employment outcomes, they are not a viable comparison group for childcare usage because their children are beyond the age when they generally require childcare. The absence of a good comparison group means that any evaluation of the impact of TFC on childcare use will carry a risk of failing to generate conclusive findings.

The recommendations for the evaluation of FEEE 30 hour extension are:

- To evaluate the ‘intention to treat’ impacts on parental employment using a difference-in-difference methodology with comparison groups of parents of children aged 1-2 and parents of children aged 5-7 applied to LFS data. This approach has a reasonable probability of identifying conclusive findings, can be undertaken at reasonably low cost and, based on current plans for the introduction and roll-out of the policy, could produce findings by 2019 (or 2020 if impacts take longer to materialise).
- To evaluate the ‘intention to treat’ impacts on childcare use using the same methodological approach applied to FRS data. Again, this approach has a reasonable probability of identifying conclusive findings, can be undertaken at reasonably low cost and, based on current plans for the introduction and roll-out of the policy, could produce findings by 2019/2020 (or 2020/2021 if impacts take longer to materialise).

In addition, the similarity in methodology and data requirements for both evaluations of parental employment impacts (and the need to consider the other policy) recommends that the evaluations would ideally be jointly undertaken.

The recommendations for the retrospective evaluations of the extension of FEEE to three year olds from 2004 and to disadvantaged two year olds in 2013/2014 are:

- For DfE to continue to monitor the findings from ongoing research considering the evaluation of the employment and childcare use impacts of these policies using LFS and FRS data and data from the Study of Early Education and Development (SEED).

Finally, the quantitative evaluation approaches could be supplemented with further information from two other sources:

Evaluation of employment impacts

- The potential impacts on parental employment and childcare use could be estimated prior to the introduction of the policies using simulations from a structural modelling approach. However, such estimates do not measure the actual impacts of the policies and are not a substitute for an evaluation. In addition, there are systematic biases which mean that these estimates may overstate the impacts and could, therefore, be regarded as ‘upper bound’ estimates of impacts.
- Qualitative research could usefully provide understanding of the mechanisms driving and reasons for the size of impacts identified in the quantitative evaluations and also inform on potential improvements in policy design.

6 Evaluation of childcare market impacts

This chapter considers the options for evaluating the impacts of TFC and FEEE on the childcare market. It first describes the required impact metrics and available data sources before assessing the evaluation options. It also examines the supplementary role of using qualitative evidence, while a final subsection makes recommendations on the best approach and summarises what it can achieve.

6.1 Impact metrics

Table 6 presents the impact metrics for an evaluation of the impacts of TFC and FEEE on the childcare market.

Table 6. Impact metrics for the childcare market

Impact metric	Description
Number of free places	Number of 15 hour and 30 hour places offered by age of child
Number of paid places	Number of paid places offered by age of child
Weekly hours per child	Average weekly hours per child by age of child
Prices	Average hourly fee by age of child
Profits	Profit as a percentage of turnover for private firms
Quality of childcare	Proxy measure for quality: distribution of staff qualifications (e.g. proportions of staff with level 3 and level 6 highest relevant qualifications)
Work facilitating features	Proxy measures for work facilitating features (e.g. offer childcare throughout the year; offer atypical hours of care; offer flexible hours).

Drawing on the logic models presented in section 3, the table presents key impact metrics for the childcare market. The first four metrics (number of places, weekly hours and prices) are relatively straight-forward to measure, both at the provider level and at the market level.

However, there is complication for the price metric. As highlighted in the logic model, TFC and FEEE may create incentives for parents to choose or providers to offer ‘better’, more costly childcare characteristics, either in dimensions which facilitate parental employment or in aspects which are beneficial to the child (including ‘quality’ related to child development). Ideally, any rise in the price measure would be divided into a ‘pure’ price rise (without any change in the

childcare characteristics) and those reflecting improvements in childcare characteristics. This requires either prices to be measured in conjunction with childcare characteristics to create ‘quality-controlled’ prices or some evidence on changes in average childcare characteristics to consider alongside observed changes in average prices. In the absence of either of these, one option is to assume that childcare characteristics do not alter substantially and that price changes primarily reflect changes in the pure price.

The final three metrics in table 6 are also challenging to measure. While profit is a well-defined concept for private firms, the notion of a ‘surplus’ for non-profit firms (i.e. a non-profit return for maintained or other non-profit settings such as being able to offer staff better conditions or pay) is problematic to measure and is not included as an impact metric.³⁴ The notion of ‘quality’ of care is also difficult to capture in a simple measure and a proxy measure of staff qualifications is offered here³⁵. Measures of work facilitating features are similarly challenging to capture in simple measures and three measures based on the offered hours are suggested here.

6.2 Data sources

Table 7 summarises the characteristics of the existing UK data sources which could potentially be used for the childcare market evaluation. Further details and reviews of other possible data sources are presented in Annex B.

Robust data is only available from existing data sources for a few of the childcare market impact metrics. A consistent annual series of the number of FEEE places is provided by the DfE, but only since 2010. The Childcare Costs Survey has collected information on prices in a consistent manner since 2002, but the methodology of data collection contains a subjective estimate of average prices at the Local Authority level. Data collection for the Consumer Price Indices (CPI) has collected information on childminder fees since 2001 and on nursery fees since 2004. The microdata (provider level) for these individual items have recently become available online (currently only from 2011, but data since 1996 will be added). The main caveat on this price data is the uncertainty, without initial analysis, of the sufficiency of the sample sizes to analyse trends over time. Other potentially useful metrics are provided in the Childcare and Early Years Providers’ Survey, but the survey is only biannual and has some inconsistencies in the data series collected.

³⁴ An alternative measure for not-for-profit providers or additional measure for private providers could be some concept of “sustainability” reflecting the provider’s perception of their ability to remain in operation given ongoing costs and income/funding levels.

³⁵ More precise quality measures, such as the ITERS, ECERS or SSTEWS measures, require specialist assessment visits to settings and could not simply be reported by providers on a regular basis in a large scale survey.

Table 7. Existing data sources for the childcare market evaluation

Data source (years)	Sample size per year	Number of places / weekly hours	Prices / profits	Quality / work facilitating features	Caveats
Family Childcare Trust Childcare Costs Survey (2002+)	Approx. 200 Local Authorities	None	Average prices for nurseries, childminders and after school care	None	LA level only Subjectivity in the reporting of prices
ONS Consumer Price Indices (currently 2011+, planned 2004+)	Circa. 500 prices for childminders and 600 prices for nursery fees	None	Prices for childminder and nursery fees	None	Uncertain sufficiency of sample sizes for analysis.
DfE Statistics on Provision for Under Fives (2010+)	All providers of early education in England	Number of FEEE places and hours	None	Staff QTS and EYPS Link to Ofsted rating	Only England Only FEEE places
DfE Childcare and Early Years' Providers' Survey (2001+)	10k providers	Number of places	Fees (only in 2013) Limited data on profits	Various quality measures Frequency & duration of services	Only England Biannual Inconsistencies in data series

Notes: QTS denotes Qualified Teacher Status and EYPS denotes Early Years Professional Status.

6.3 Options for TFC and the FEEE 30 hour extension

The options for evaluating impacts on the childcare market are limited by the lack of a comparison group. Although the FEEE 30 hour extension applies only to three and four year old children, the logic model in section 3 highlighted how providers may adjust their provision for all ages of children in response to offering the free places to the targeted age group. Unlike parental choices over work and childcare, such 'spillover' effects to other ages are likely to be of some importance for the childcare market metrics. Hence, trends for non-eligible ages

Evaluation of childcare market impacts

of children cannot be used as a comparison group for an evaluation of the impacts of the FEEE 30 hour extension on the childcare market.

The combination of an absence of a comparison group for either policy and the concurrent introduction of TFC and the FEEE 30 hour extension mean that the impacts of the two policies on the childcare market cannot be separately identified.

There is only one option for the evaluation of the joint impact of the introductions of TFC and the FEEE 30 hour extension:

- A simple discontinuity in time trends approach to identify whether the policy introductions coincide with a distinct break in ongoing trends which cannot be attributed to any other factors.

6.3.1 Discontinuity in time trends

a) What question is being answered?

The question being addressed is whether the introduction of TFC and the FEEE 30 hour extension together have an impact on the childcare market metrics.

b) What is the evaluation methodology?

The evaluation methodology compares prior trends in the childcare market metrics with the changes that occur over the period when the two policies are introduced and statistically tests whether and to what degree the changes are different to prior trends. The regression analysis could also include control variables for other known influences on the childcare market variables³⁶.

c) How robust is the approach?

The robustness of the approach is dependent upon the strength of evidence that any discrete change in the childcare metrics cannot be explained by another influence or concurrent change in a different factor. Hence, robustness is challenging to positively prove as it depends upon an absence of contrary evidence.

d) What is the best data source to use?

Ideally, this approach would use data providing a long, consistent historic series with which to estimate 'normal' trends. The only existing data sources which potentially offer this are the Childcare Costs Survey or CPI data for the childcare price metric. However, the subjective element in the collection of the Cost

³⁶ For example, this could include the population size of children in the age cohort in the analysis of the number of places or cyclical economic trends (such as GDP growth) in the analysis of profit rates.

Survey data means that it is only indicative of price trends over time. One drawback of the CPI data is that it is only currently available from 2011 (although due to become available since 2004) and it is not clear, without initial analysis of the variation in prices, what size of discontinuities in time trends that the moderate sample sizes are sufficient to detect. In addition, the CPI source provides no other information on the childcare market metrics, including any indication whether price changes reflect changes in the characteristics of childcare. Nevertheless, the CPI data could provide an additional important source of indicative information on childcare price trends, particularly when the earlier years of data become available.

An alternative option is a bespoke survey of providers undertaken in 2016 and 2018 (and ideally later) collecting information on the childcare market impact metrics. However, in order to estimate longer term trends leading up to 2016, the initial survey would need to collect retrospective information on the metrics. While the collection of retrospective data can be challenging and the data suffer from potential recall bias, these issues may be mitigated by the fact that the data would be collected from organisations (rather than individuals) which may have formal records of previous provision.

The sample size would need to be sufficiently large in order that the estimates of previous trends can be adequately precise to allow the identification of a discrete break in the trends. Estimates of the required sample size would entail information on the degree of variation in the impact metrics which is not available. Given the relative simplicity of the impact metrics, it may be possible for the data to be collected using an online survey which would reduce the survey costs.

Another option could be the incorporation of the collection of the required data into the ongoing Childcare and Early Years Providers Survey. It is unlikely this could be considered for the next survey for which fieldwork is to be undertaken in March to June 2016, but the collection of retrospective data in the following survey, likely to be in 2018, could be considered.

e) What are the time and monetary costs involved?

Based on current plans for the introduction and roll-out of both policies, findings from a bespoke survey could be available in 2019, although subsequent surveys and analysis would be beneficial to consider trends after the introduction of the TFC and the FEEE 30 hour extension. If face-to-face interviews were required for the bespoke survey, the cost of the primary data collection could be substantial: for a sample of 5,000 providers, the costs could be in the region of £1 million, although an online survey would have a considerably lower cost.

Evaluation of childcare market impacts

f) What is the overall assessment of the approach?

This approach is likely to identify descriptive statistics of trends in the childcare market metrics which could be consistent with joint impacts of TFC and the FEEE 30 hour extension, but could not provide a robust evaluation of impact for either individual policy or for the joint impact of both policies.

6.4 Options for FEEE: previous extensions

As highlighted in section 6.2 above, there is no existing source of a robust and consistent series of data on the childcare market impact metrics covering the previous extension to the FEEE from 2004 (and only limited data covering 2013/14). In addition, the extension to three year olds which began in 2004 was gradually implemented over a number of years so that any impact is unlikely to be identifiable as a distinct break in ongoing trends. Retrospective collection of data over such a long period is unlikely to be feasible. Hence, evaluations of the impacts of the previous FEEE extensions on the childcare market are not feasible due to a lack of adequate data sources.

6.5 Role of qualitative evidence

As was the case with the impacts on parental employment, the collection of qualitative evidence on the impacts of TFC or the FEEE 30 hour extension on the childcare market could both provide supplementary evidence on the metrics considered in the quantitative evaluation approaches and also compliment the quantitative approaches by providing information to help understanding of the reasons and mechanisms for the findings from the quantitative approaches.

For the first type of information, a qualitative approach might ask providers how they thought the introduction of TFC or the FEEE 30 hour extension had affected their childcare provision, including the number and type of places, prices and profitability. Separate identification of the impacts from each policy by providers might be facilitated by the fact that they would be reacting to different sources of change for each policy, that is, to changes in parental demand in response to TFC and to the funding offer from Local Authorities in the response to the FEEE 30 hour extension. Indeed, the reaction to the FEEE 30 hour extension could be quite a distinct point of decision, although it might be more challenging for providers to disentangle the effects of TFC from other ongoing trends in childcare demand from parents. The second type of information on the understanding of mechanisms and reasons for particular impacts might also be more useful for the FEEE 30 hour extension than TFC. Again, the decisions surrounding responses to FEEE 30 hour extension are more likely to have been made at a clear, discrete point and the underlying reasoning may be more salient to providers. In addition, the complexity and variety of business models used by

childcare providers would make a qualitative, less structured approach particularly valuable to collect this type of information.

Qualitative research with providers would therefore be most usefully undertaken for the FEEE 30 hour extension, although information about TFC could be collected as well. It would not require a large sample, although the sampling strategy should ensure that a broad range of different types of childcare providers are included to fully capture different motivation and contexts that might influence responses to the policies (for example, a range across private and maintained providers; chains and single site providers; and urban and rural providers).

6.6 Recommendations

The options to evaluate the impacts of TFC and the FEEE 30 hour extension on the childcare market are severely limited by the lack of a comparison group for either policy; by the concurrent introduction of both policies; and by the absence of robust, consistently collected data on the childcare market impact metrics.

The recommendations for the evaluation of both policies are:

- A simple discontinuity in time trends approach using primary data collection to identify whether the policy introductions coincide with a distinct break in ongoing trends which cannot be attributed to any other factors. However, this approach has substantial drawbacks: the primary data collection would be expensive; the analysis would need to rely on data that had been collected retrospectively; and the methodology identifies indicative descriptive statistics which may be consistent with hypotheses of the joint impact of both policies but could not provide a robust evaluation of impact for either individual policy or for the joint impact of both policies.
- A qualitative approach using interviews with providers to help identify their responses to the FEEE 30 hour extension (and any views on the impact of TFC on their provision).

Retrospective evaluations of the impacts of the previous FEEE extensions on the childcare market are not feasible due to a lack of adequate data sources.

With respect to the primary collection of quantitative data on the childcare impact metrics from providers, it is worth noting that a more frequent (annual) Childcare and Early Years Provider's Survey with a focus on the collection of a consistent series of key childcare market metrics from the entire sample would have been particularly useful to the evaluations considered here. Indeed, given the ongoing development of early education and childcare policy and the likelihood that understanding the impacts of policy on the childcare market will

Evaluation of childcare market impacts

continue to grow in importance, further investment in this survey to develop a more regular and robust source of information is recommended.

7 Conclusion

Identifying a robust methodology to evaluate policy is often challenging and heavily dependent upon policy design and the contextual opportunities for finding a suitable comparison group to estimate counterfactual outcomes. Feasibility of an approach is often more of a question of the likelihood of obtaining conclusive findings rather than whether estimation of an impact can be attempted. In addition, the risks of failure to answer the evaluation questions with a particular approach often depend upon unknown parameters such as the size of impact that the evaluation is seeking to estimate. Nevertheless, some clear conclusions can be reached on the feasibility of evaluating the impacts of TFC and the FEEE on parental employment and the childcare market.

Evaluation of the employment impacts is facilitated by the availability of comparison groups of parents whose children are not of eligible age for the policies and by an existing, large data source with the required impact metrics. However, the evaluation of the impact of TFC on childcare use is seriously challenged by the lack of a comparison group as most parents who might need to use childcare are potentially eligible for the policy. The analogous evaluation for the FEEE does not face the same challenge as the age-specific aspect of the policy means that it excludes a substantial proportion of children for whom parents are likely to require childcare. The options to evaluate the impacts on the childcare market for either policy are severely limited by the lack of any comparison group; the concurrent introduction of the policies; and by the absence of robust, consistently collected data on the childcare market impact metrics.

Annex A: Literature review of childcare policy evaluations

Table 8. Summary of evaluations: UK studies

Study and data	Methodology	Outcomes	Main results	Caveats
Blanden et al (2014) LFS 2002-2007 & DfE data on availability of places	Difference-in-differences using speed of implementation across geographical areas to evaluate the impact of introduction of FEEE in England for three year olds in 2004.	Employment of mothers with three years olds. Proportion of three year old children in formal childcare	No effect on mothers' employment. Positive effect on childcare use, but partially offset by crowding out of paid places.	Speed of implementation endogenous to child-age-specific employment outcomes. Working paper.
Blundell, Duncan, McCrae & Meghir (2000) FRS 1994-1996	Structural modelling: discrete behavioural model of labour supply with imputed childcare costs for non-working parents using distribution of childcare expenditure within household type to allow for heterogeneity in quality.	Employment and work hours for married and single mothers and married fathers	Effects of childcare cost not directly presented.	Response to childcare subsidies may not be the same as to lower childcare prices.
Brewer, Cattan, Crawford & Rabe (2014) LFS 2000-2013 & DfE data on availability of places	Two methods: (a) regression discontinuities design using eligibility cut-offs with mother fixed effects and (b) difference-in-differences using speed of implementation across geographical areas across ages of children to evaluate the impact of introduction of FEEE in England for three year olds in 2004	Employment of mothers with three year-olds. Also for those with no younger children. Also for part/full-time employment	Positive effects on mothers' employment of 2.8 and 6.3 percentage points ((a) and (b)). Impact greatest for those with no younger children. Impact greatest for movement into part-time work.	(a) May understate impact due to anticipatory effects or delayed responses. (b) May overstate impact if speed of implementation endogenous to child-age-specific employment outcomes.
Brewer & Crawford (2010) Administrative data 1999-2007	Regression discontinuities design using eligibility cut-offs and placebo tests with children of other ages to evaluate the impact of eligibility for FEEE at age 3 and compulsory schooling at age 4.	Employment and welfare receipt for lone parents receiving welfare	Small effect on employment and welfare receipt for compulsory schooling and weak effects for FEEE.	May understate impact due to anticipatory effects or delayed responses. Working paper.
Jenkins & Symons (2001) 1989 Lone Parents Survey	Structural modelling: probit model for employment using estimated net wage and childcare price (from observed cost) to estimate the impact of childcare price.	Employment of lone mothers	Small negative impact of childcare costs on employment.	Older data when employment rate lower. Estimated price does not control for quality. Response to childcare subsidies may not be the same as to lower childcare prices.

Paull, Taylor & Duncan (2002) FRS 1994-1999	Structural modelling: discrete 5-state behavioural model with childcare costs using standardised-quality prices (from variation across Local Authorities) to estimate the impact of childcare price.	Use of formal childcare, weekly hours of childcare and hourly amount spent on childcare	Negative impact of childcare price on use of formal childcare for preschool children; no impact on weekly childcare hours; and negative impact on amount spent per hour.	Response to childcare subsidies may not be the same as to lower childcare prices.
Viitanen (2005) FRS 1997-2004	Structural modelling: bivariate probit model for employment and childcare using estimated wage and childcare price (from observed cost) to estimate the impact of childcare price.	Employment of married mothers of pre-school children and use of childcare.	Positive effect of subsidies on mothers' employment and childcare use.	Estimated price does not control for quality. Response to childcare subsidies may not be the same as to lower childcare prices.

Table 9. Summary of evaluations: other countries

Study and data	Methodology	Outcomes	Main results	Caveats
Berlinski & Galiani (2007) Argentina: Encuesta Permanente de Hogares (household survey) and data on pre-school expansion 1992-2000	Difference-in-differences approach using differences in speed of the large expansion in free pre-school across geographic areas between 1994 and 2000. Regional fixed effects included to address the possibility that the speed of the construction of new facilities may have depended upon initial differences in enrolment rates.	Employment and hours worked for mothers with a child aged 3-5 years old. Pre-primary school use by 3-5 year olds. For mothers with/without spouses. For those with/ without younger children	Positive effects on pre-primary school use and maternal employment. Similar results regardless of spouse or younger children. No effect identified for hours of work.	
Berlinski, Galiani & McEwan (2011) Argentina: Encuesta Permanente de Hogares (household survey) 1995-2001	Regression discontinuities design using eligibility cut-offs to evaluate the impact of eligibility for compulsory pre-school at age 4.	Employment for mothers with a child aged 4, with and without younger children.	Positive effects on employment rates and work hours for mothers without younger children. No effects for mothers with younger children.	Children may attend pre-school prior to this year (only small number) or may not attend in spite of being compulsory.
Farfan-Portet, Lorant & Petrella (2011) Belgium: register of tax claims and data on number of childcare places 1994-2001	Logit models with year dummies for the years that childcare subsidies changed and control variables for childcare availability at the municipality level. Time trend also included.	Use of paid childcare. For different income groups.	Positive effects on use of paid childcare. Effect of increasing availability greatest for low income families. Effects of tax policies across income groups depended on the nature of the change.	Availability of places could be endogenous to demand. Year dummy variables may be capturing other coincidental changes.

Annex A: Literature review of childcare policy evaluations

<p>Baker, Gruber & Milligan (2008)</p> <p>Canada: National Longitudinal Survey of Children and Youth (NLSCY) 1994-2003</p>	<p>Difference-in-differences comparing Quebec with the rest of Canada before and after the introduction of \$5 a day childcare subsidy.</p> <p>Evidence suggesting that policy not endogenous to outcomes. Lone mothers excluded due to concurrent other policy reforms. Checks for Quebec-specific shocks using trends before and after treatment; comparison with 8-11 year-old; and economic control variables.</p>	<p>Childcare use by 0-4 year-olds.</p> <p>Employment of women in two-parent families.</p>	<p>Positive effects on childcare use and maternal employment.</p>
<p>Lefebvre & Merrigan (2008)</p> <p>Canada: Survey of Labour and Income Dynamics 1993-2002</p>	<p>Difference-in-differences comparing Quebec with the rest of Canada before and after the introduction of \$5 a day childcare subsidy across different years.</p>	<p>Employment, annual weeks worked, annual hours worked and annual earnings for mothers with a child under age 6.</p> <p>Also by education level.</p>	<p>Positive effects on employment, hours worked, weeks worked and earnings. Effects increase over the years. Effects (other than earnings) greater for less educated.</p> <p>Could be concurrent Quebec-specific shocks on outcomes.</p>
<p>Lefebvre, Merrigan and Verstraete (2011)</p> <p>Canada: Survey of Labour and Income Dynamics 1996-2004</p>	<p>Difference-in-differences comparing Quebec with the rest of Canada before and after the introduction of \$5 a day childcare subsidy across different years.</p> <p>Difference-in-difference-in-differences (DiDiD) with additional control groups of mothers with children only aged 12-17.</p> <p>Evidence suggesting policy is not endogenous to outcomes.</p>	<p>Employment, annual weeks worked, annual hours worked and annual earnings for mothers only with children aged 6-11.</p> <p>Also by education level.</p>	<p>Positive longer term effects (child aged 6-11) on employment, hours worked, weeks worked only for less educated mothers.</p>
<p>Goux & Maurin (2010)</p> <p>France: Census data for March 1999</p>	<p>Regression discontinuities design using eligibility cut-offs to evaluate the impact of free pre-school at ages 2 and 3.</p>	<p>Employment for mothers with children aged 2 and 3.</p> <p>Also by education level.</p>	<p>Positive effects for single, less educated mothers, particularly at age 3, but no effects for single more highly educated mothers or two parent families.</p>
<p>Bauernschuster & Schlotter (2015)</p> <p>Germany: Socio-Economic Panel 1991-2005 and Micro-Census from 1991,1993, 1995, 1996 and 2001</p>	<p>Two approaches: (a) Regression discontinuities design with instrumental variables to estimate geographical variation in eligibility cut-offs and (b) Difference-in-differences with mothers with youngest children aged 10-11 and women without children to evaluate the introduction of highly subsidized kindergarten from age 3 to 6 in 1996. Placebo test in pre-treatment period to check trends same for the treatment and control groups</p>	<p>(a) Employment and work hours for mothers with youngest children aged 3 and 4.</p> <p>(b) Employment for mothers with youngest children aged 3 and 4.</p>	<p>(a) Positive effects of eligibility on mothers' employment and work hours.</p> <p>(b) Positive effect on mothers' employment. Larger effects for single mothers than mothers with partners.</p>
<p>Haan & Wrohlich (2011)</p> <p>Germany: Socio-Economic Panel (SOEP) 2000-2007</p>	<p>Structural model of female employment and fertility using variation in financial incentives from the tax-transfer system and regional variability in cost and availability of subsidized places to simulate the impacts of increased availability of subsidised places for 3 year olds with two working parents.</p>	<p>Employment and fertility for married and cohabiting women</p>	<p>Positive effects on employment.</p> <p>Positive effect on fertility only for highly educated women and women without children.</p>

Annex A: Literature review of childcare policy evaluations

Schlosser (2011) Israel: Labour Force Survey 1998-2003	Difference-in-differences using geographical variation in the speed of implementation to evaluate the introduction of free pre-school for 3 and 4 year olds in 1999. Triple difference models using mothers with children of other ages (0-1 and 5-9). Placebo test in pre-treatment year.	Labour force participation, whether studies or works, employment rate and work hours for Arab mothers with children aged 2-4. Also mothers with youngest child aged 3-4.	Positive effects on all labour force outcomes Greater employment effects for mothers without younger children. Effects only for more educated mothers.	Working paper.
Bettendorf, Jongen & Muller (2012) Netherlands: Labour Force Survey of Statistics 1995-2009	Difference-in-differences (DiD) using parents with older children (age 12-17) as a control group to evaluate an increase in childcare subsidies during 2005-2009. Employment trends similar for treatment and control groups before treatment. Test a placebo effect during pre-treatment period, although not passed for subgroups.	Employment and work hours for mothers and fathers with a child under 12.	Positive effects on employment and work hours for women. Greater effect for more highly educated women, single women over women in couples, for women with younger children and for immigrant over native women. No effect on employment for men, but small decline in work hours.	Effects are upper bounds due to concurrent rise in the earned income tax credit for the treatment group. Working paper.
Havnes & Mogstad (2011) Norway: Administrative Registers and data on childcare coverage by municipality 1967-2006	Difference-in-differences (DiD) using geographical variation in speed of expansion across municipalities to evaluate large-scale expansion of subsidised childcare for children aged 3-6 from 1975. Placebo test in pre-treatment period for differential trends between treatment and comparison group. Triple differences (DiDiD) models of mothers with youngest child aged 7-10. Include pre-reform municipality variables.	Employment for married mothers with children aged 3-6.	Little or no effect on maternal employment because formal care crowded out informal care.	Speed of implementation may be endogenous to outcomes.
Nollenberger & Rodriguez-Planas (2015) Spain: Spanish Labour Force Survey 1987-1997	Triple difference (DiDiD) using geographic variation in speed of implementation and comparison with mothers with younger children (2 years old) to evaluate the introduction of subsidised childcare for all 3 year olds from the early 1990s. Include separate time trends for treatment and comparison groups. Placebo tests in pre-reform period. Tests using comparisons with older children. Exogeneity of speed of reform tested using reform dummy for one year prior to reform. DiD approach without geographical variation generated similar results.	Employment and work hours for mothers with 3 year old children in short run (at age 3) and long run (at age 7).	Positive effect on employment and work hours in short run which persist in long run. Stronger for less educated mothers.	Discussion paper

Annex A: Literature review of childcare policy evaluations

<p>Gustafsson & Stafford (1992)</p> <p>Sweden: National Household Survey and data on fees and availability 1984</p>	<p>Logit and ordered probit models using variation in childcare fee structure and availability of places across municipal governments in 1984 to estimate the impact of childcare price and subsidies.</p>	<p>Employment for mothers and childcare use.</p>	<p>Negative effect of price on use and mothers' employment.</p>	<p>Small sample. Fee variation could be endogenous to the outcomes.</p>
<p>Lundin, Mork & Ockert (2008)</p> <p>Sweden: Register Data and survey data on childcare prices for 2001 and 2003</p>	<p>Difference-in-difference matching regression using variation in childcare price reductions across family type and region (measured in household type and municipality fixed effects) to evaluate the introduction of a cap on childcare prices in 2002.</p>	<p>Employment and share in full-time employment for mothers in two-parent households with a child aged 1-9.</p>	<p>No effects.</p>	<p>Employment rates already high.</p>
<p>Felfe, Lechner & Thiermann (2013)</p> <p>Switzerland: Strukturhebung and municipality data 2010</p>	<p>Matched sample approach with analysis restricted to region borders to address endogeneity of policy to evaluate geographical variation in regulations of after-school care provision.</p>	<p>Employment and work hours for mothers and fathers with children aged 0-12 years.</p>	<p>No effect on employment for mothers or fathers. Positive impact on proportion working full-time for mothers and negative effect for fathers.</p>	<p>Discussion paper</p>
<p>Bassok, Fitzpatrick & Loeb (2014)</p> <p>USA: Longitudinal Business Database (private provision) & Common Core of Data (public provision) 1990-2010</p>	<p>Difference-in-differences using geographical variation across states in preschool provision to evaluate the introduction of a voucher program in Georgia (from 1995) and expansion of public provision in Oklahoma (from 1998). Control groups are (i) all other states (ii) other southern states and (iii) a synthetic control group, weighting other states on the basis of similar characteristics to the treatment group prior to treatment.</p>	<p>Number of private and public childcare centres and their employment. Pay per employee as a proxy for quality.</p>	<p>Vouchers had positive effects on both number of private and public providers. Expansion of public provision shifted workers to public provision, but no overall increase. Consistent with hypothesis that government funding of a subsidy expands a market more than government provision.</p>	<p>Assume that changes in Georgia and Oklahoma were independent of any other changes in the childcare sector in these states.</p>
<p>Cascio (2009)</p> <p>USA: Decennial Census data from 1950 to 1990 and data on first year of kindergarten funding</p>	<p>Triple difference model using geographical variation in expansion of kindergarten for 5 year-olds in comparison to mothers with slightly older and slightly younger children to evaluate the expansion in kindergarten for 5 year olds from mid-1960s to late 1970s. Restriction to states with expansion improves matching. Evidence suggesting policy not endogenous to outcomes and long period of pre-treatment data avoids any pre-treatment dip.</p>	<p>Employment and use of kindergarten for single and married mothers with 5 year old children With and without younger children.</p>	<p>Positive effect on employment only for single mothers without younger children. Shift from private to kindergarten provision for all groups except single mothers with younger children.</p>	

Annex A: Literature review of childcare policy evaluations

Cascio & Whitmore Schanzenbach (2013) USA: Current Population Survey 1977-2011	Triple differences model using geographical variation across states in preschool provision and in comparison to mothers of 5 year olds to evaluate the introduction of a voucher program in Georgia (from 1995) and expansion of public provision in Oklahoma (from 1998). Control group is all other states.	Employment for mothers of 4 year olds. Pre-school enrolment at age 4 (for DiD model). For single/married mothers and for those with/without younger children	Positive effect on employment for less educated mothers only. Positive effect on pre-school enrolment. Greater impact for mothers with lower education. Effects consistent across subgroups.	
Fitzpatrick (2010) USA: Decennial Census 2010	Regression discontinuities design using eligibility cut-offs in comparison to states without universal kindergarten to evaluate the impact of free kindergarten for 4 year-olds in Georgia and Oklahoma. Need comparison group because enrolment increases at age 4 in the absence of a program.	Employment for mothers of 4 year olds. Pre-school enrolment at age 4	No effect on employment. Positive effect on pre-school enrolment.	
Fitzpatrick (2012) USA Decennial Census 2010 and data on kindergarten cutoff dates by state	Regression discontinuity instrumental variables framework (RDIV) using geographical variation in eligibility cut-offs for free kindergarten as an instrument for enrolment to evaluate the impact of public school enrolment on maternal employment.	Employment for mothers of 5 year olds by single/married and with/without younger children.	Positive effect on single mothers without young children. No effect for other mothers.	
Gelbach (2002) USA: Census 1980	Regression discontinuity instrumental variables framework (RDIV) using quarter of birth (related to geographical variation in eligibility cut-offs) to instrument for enrolment to evaluate public school enrolment for 5 year olds.	Employment for mothers of 5 year olds by single/married and with/without younger children.	Positive effects for married mothers with and without younger children and single mothers without younger children. No effect for single mothers with younger children.	Requires that maternal employment not affected by the age of child other than through enrolment in public school.
Guner, Kaygusuz & Ventura (2013) USA: calibrations from various sources	Theoretical life-cycle equilibrium model of household labour supply using calibrated parameters from various data sources to estimate the impact of more generous childcare subsidies.	Male and female employment and hours of work.	Positive effect on female participation and work hours and negative effect on male work hours.	Theoretical model. Discussion paper.
Ryan, Johnson, Rigby & Brooks-Gunn (2011) USA: Fragile Families and Child Wellbeing Study 2002/2003	Propensity score matching to identify non-recipient comparison groups for subsidy recipients to evaluate the use of childcare subsidies in the Child Care and Development Fund.	Quality of care for 3 year-olds.	Positive impact on quality choices through being more likely to choose centre-based care.	Relatively small sample.

Annex A: Literature review of childcare policy evaluations

<p>Weinraub, Shlay, Harmon & Tran (2005) USA: small scale bespoke survey (111)</p>	<p>Tests of statistical significance for differences between treatment and control group to evaluate the impact of receipt of any type of childcare subsidy.</p>	<p>Quality of care used by low income African American families with a child under 5</p>	<p>Positive effect on use of centre and formal care (relative to other types), but no effect on quality.</p>	<p>Very small sample.</p>
<p>Marrufo, O'Brien-Strain & Oliver (2003) Davis & Li (2005) Davis, Li, Weber & Grobe (2009) USA: Market rate surveys in multiple years</p>	<p>OLS regression of county-level childcare prices in California, Minnesota and Oregon on factors influencing supply and demand and childcare subsidy expenditures in panel models with random county effects to estimate the impact of subsidies.</p>	<p>County-level average childcare price.</p>	<p>Positive effect only in California</p>	<p>Small sample size and few statistically significant coefficients. Prices may vary within counties as much as between counties. No causal links can be identified.</p>

Annex B: UK data sources

Table 10. UK data sources: labour market / childcare usage

Source	Annual sample, frequency, survey type	Parental employment	Childcare Details	Use of Policy	Controls	Availability
Labour Force Survey (LFS)	300k adults (55k mothers, 45k fathers) in UK Quarterly since 1992 Quasi-panel	Work status / hours / earnings Other work characteristics Reasons for work choices	Type used for each child (autumn quarters only from 2001 to 2009)	Receipt of CTC and WTC from 2001	Broad range of child, adult and household variables Income by sources	Available at UK data archive (ONS)
Childcare and Early Years Survey of Parents (+ previous versions)	6.5k families (6.5k mothers, 5k fathers) with children under age 15 in England Occasional since 1999 Cross section	Work status / hours Reasons for work choices	Hours and payment by child and type Reasons for usage / why might start using Views on quality, availability, affordability of main provider	Receipt and amount of CTC and WTC Use of ESC Number of FEEE hours for each child	Household income Income sources for those WTC eligible (no amounts) Household composition, education, marital status, ethnicity.	Available at UK data archive (DfE) Data series not consistent over time.
Family Resources Survey (FRS)	35k adults (6k mothers, 4.5k fathers) in Britain Annual since 1994 Cross section	Work status / hours / earnings	Type, hours, and payments for each child	Receipt and amount of CTC and WTC. Use of CCTC and ESC	Broad range of child, parent and household variables Income by sources	Available at UK data archive (DWP)
Family Expenditure Survey (FES) / Expenditure and Food Survey (EFS) / Living Costs and Food Survey (LCFS)	5k – 7k households (2K mothers, 1.5k fathers) in UK Annual from 1961 Cross section	Work status / hours / earnings	Whether attends pre-school for each child. Payment amounts for “nursery, crèche and playschool” and childcare payments for all children	Receipt and amount for CTC and WTC	Broad range of child, adult and household variables Income by sources Household consumption	Available at the UK Data Archive (ONS)
HMRC Tax Credits Database	Administrative data at award level	Eligibility for basic + 30 hour element of WTC. Whether in receipt of Income Support/JSA in household	Daily value of eligible childcare costs	WTC amount including CCTC amount.	Household composition. Household income LA and region	Restricted availability at HMRC data lab (08/9 - 10/11)
Families and Children Study (FACS)	7k families with children (7k mothers, 5k fathers) in Britain Annual 2001-2008 Panel (8 waves)	Work status / hours / earnings Other work characteristics Reasons for work choices	Hours and payment by child and type Reasons for usage Whether price changed in past year and effect on usage	Receipt and amount of CTC and WTC	Broad range of child, parent and household variables Income by sources	Available at UK data archive (DWP)

Annex B: UK data sources

British Household Panel Survey (BHPS)	10k adults households (1.5k mothers, 1k fathers) in UK Annual 1991-2009 Panel (18 waves)	Work status / hours / earnings Other work characteristics	Types and amount paid for all children for working parents	Receipt of CTC WTC	Broad range of child, adult and household variables Income by sources	Available at UK data archive (ONS)
Understanding Society	50k adults (10k mothers), 8k fathers) in UK Annual since 2009 Panel (4 waves)	Work status / hours / earnings Other work characteristics	Type and hours for each child	Receipt and amount of CTC and WTC	Broad range of child, adult and household variables Income by sources	Available at UK data archive
DfE Study of Early Education and Development (SEED)	5k households with a child aged 2 in England Once 2013-2016 Panel (3 waves) at ages 2, 3 and 4	Work status / hours Reasons for work choices	Weekly hours and duration of use by type for target child only.	Take-up of FEEE for target child only.	Broad range of child, adult and household variables	Not publically available

Notes: Reports on the Childcare and Early Years Survey include La Valle et al (2000), Woodland et al (2002), Bryson et al (2006), Kazimirski et al (2008), Smith et al (2010), Smith et al (2012), Huskinson et al (2013), and Huskinson et al (2014). Surveys which were not relevant: Annual Survey of Hours and Earnings (ASHE) (no childcare information); National Child Development Study (NCDS) (1958 cohort too old).

Table 11. UK data sources: childcare markets (providers)

Source	Sample, frequency, type	Quantity/Provider Type	Price / Cost / Fees & Business	Provision of Entitlement	Quality	Availability / caveats on consistency
Childcare Costs Survey (Family and Childcare Trust)	Most (~=200) local authorities in Britain Annual since 2002 Repeated cross section	Whether have sufficient childcare coverage based on the latest childcare sufficiency assessments.	Average fees for 25 and 50 weekly hours by type (nurseries, childminders, after school care). Fees estimated by Family Information Services in LAs	Proportion of eligible children receiving FEEE	None	Statistics published in reports published by Family and Childcare Trust (see Daycare Trust (2008, 2009, 2010, 2011, 2012, 2013), Rutter and Stocker (2014), Rutter (2015))
ONS Consumer Price Indices	Quarterly price data on circa. 150 nursery fess and 130 childminder fees for children aged 0-4	None	Childminder fees (2001–ongoing), nursery fees (2004–ongoing), after school club charges (2004–2013)	None	None	Data from 2011 currently available but planned to be available from 1996. Available at: http://www.ons.gov.uk/ons/gui-de-method/user-guidance/prices/cpi-and-rpi/cpi-and-rpi-item-indices-and-price-quotes/index.html .

DfE Statistics on Provision for Under Fives (from Early Year Census and School Census)	All providers of early education in maintained, private, voluntary and independent settings in England Annual since 2010 Cross section	Number of FEEE places and attendees for children under age 5 by provider type, region, and local authority	None	Use by funded / non-funded children by age and provider type. Number of funded hours by age and type.	Whether staff have QTS and EYPS. Can be linked to Ofsted rating.	Reports available at ONS website (see Department for Education (2014)) LA level data available
Childcare and Early Years Providers' Survey	10k PVI and maintained group-based providers and childminders in England. Since 2001, annual since 2005 Cross section	Provider type, number of places, number of children Frequency/duration of services	Fees by banded ages of children. Whether fees increased / vary across children. Whether provider is profitable or grew.	Whether offers FEEE Number of children on FEEE Number of children with fees paid by employer vouchers or local authorities	Qualifications / training of staff Staff: child ratios Average hourly pay of staff Retention of staff	Available at the UK Data Archive (DfE) See Brind et al (2012a, 2014) Inconsistencies in data series and limited availability of earlier data
Childcare Provider Finances Survey	1k+ childcare and child-minders (follow-up to providers survey) Once, 2012 Cross section	Provider type, number of children, type of care provided	Average fees + change over 2 years Costs by type; income by source; profitability; investment needs, financial concerns	Amount of income by source. Share of children paying by parental employer or voucher	None	Available at the UK Data Archive (DfE) See Brind, Norden and Oseman (2012b)
Children's Nurseries: UK Market Report (Laing and Buisson)	Up to 15 percent sample of nurseries offering full-day care in UK Annual since 2002	Market research: sector value, income sources, occupancy rates, number of places attendance, parental employment.	Average full day fees by sector and region	None	Staff pay rates	Private market research report available only by purchase.
NDNA Annual Nursery Survey	Annual survey of circa. 300 nurseries (PVI (99%) and maintained (1%))	Type of provider	Hourly funding and hourly loss per child for 2 and 3/4 year olds	Whether offering FEEE for 2 year olds and 3/4 year olds	Ofsted rating	Statistics published in report by NDNA (see NDNA (2015))
DfE Study of Early Education and Development (SEED)	Visits to 160 settings with provision for 2 or 3 year olds Once, 2015 Cross section	Type of provider	Hourly delivery cost by banded ages of children (under 2, 2, 3-4 pre-school and 4+ school)	Number of children and hours receiving FEEE.	Linked to ITERS, ECERS and SSTEW quality measures.	Not publicly available

Annex B: UK data sources

CEEDA / Playschool Learning Alliance	Diary data collection from 100 PVI nurseries with good or outstanding Ofsted rating in England Once, 2014 Cross section	Type of provider	Hourly delivery cost for funded two, three and four year olds.	Numbers of entitled children	Ofsted rating	Published report (Ceeda (2014))
--------------------------------------	--	------------------	--	------------------------------	---------------	---------------------------------

Table 12. UK data sources: childcare markets (business data)

Source	Sample, size, frequency	Data	Childcare identifiers	Availability
Inter-Departmental Business Register (IDBR)	Administrative data on all UK businesses under the VAT and PAYE schemes. Continually updated: from 1997	Employment, turnover, year of birth/death of enterprise/local unit, reason for death of local unit, legal status, regional codes (GOR, County, TTWA, Constituency, SOA)	5-digit SIC codes	Data are not available outside Government. 2-digit SIC code statistics available in Business Population Estimates reports
Business Structure Database (BSD)	Annual snapshots of IDBR. Panel from 1997	As above	5-digit SIC codes	Restricted use micro-data at VML.
Annual Business Survey (ABS)	73k survey of businesses from IDBR Annual from 2009 (links to Annual Respondents Database 1997-2008) Cross section	Gross value added, value of purchases, employment costs, net capital expenditure, turnover, banded employment and region.	5-digit SIC codes	Aggregate statistics available at ONS for 4-digit SIC codes Restricted use micro-data at VML for 5-digit SIC code
Business Register and Employment Survey (BRES)	80k survey of businesses from IDBR Annual from 2009 (links to Annual Respondents Database 1997-2008) Cross section	Number of employees (full- and part-time), country, region and Local Authority.	5-digit SIC codes	Aggregate statistics available at ONS. Restricted use micro-data at VML for 5-digit SIC code

Notes: Relevant SIC codes: 85100: Pre-primary education (pre-primary education is defined as the initial stage of organised instruction designed primarily to introduce very young children to a school-type environment, that is, to provide a bridge between the home and a school-based atmosphere) and 88910: Child day-care activities (activities of day nurseries for pupils, including day-care activities for disabled children)

Annex C: Power calculations

This annex presents the power calculations undertaken to estimate the minimum impact sizes that might be detected in the evaluations of TFC and the FEEE 30 hour extension.

Power calculations are a method of estimating the minimum size of impact that may be detected from an evaluation approach using a specific data source. This minimum size is typically termed the ‘minimum detectable effect’ or MDE. The estimated MDE depends upon the sample sizes in both the group subject to the policy and the comparison group and on the degree of variation in the impact metric. In general, larger sample sizes or less variation in the impact metric increase the precision (degree of accuracy) with which the metric is measured for both groups and increases the likelihood that smaller impacts can be identified as being statistically significant.

It should be noted that MDE estimates are calculated with a degree of uncertainty³⁷ and that they require information on parameters which are themselves only informed estimates. In the estimates calculated here, these parameters include estimates of sample sizes based on current numbers in the surveys considered and estimates of the means and variance of the impact metrics based on currently observed levels for the comparison groups. In addition, the difference-in-difference approaches inherent in the methodologies considered here create additional issues for the calculation of MDEs that cannot be fully addressed using the standard statistical software currently available³⁸. As an approximation, the MDEs are estimated using a simplified approach of sample sizes during the ‘treated’ period for the two groups with current parameters on impact metrics for the comparison group.³⁹

In order to simplify the data preparation (and not require special access data with dates of birth), all three year old children have been used as a proxy for the childcare use measure and sample sizes for three and four year old preschool children who are eligible for the FEEE 30 hour extension. This may slightly understate the sample sizes by omitting all the eligible four year olds and

³⁷ To be precise, MDEs include risks of falsely accepting that there is an impact (a type I error) and falsely rejecting that there is an impact (a type II error). In the power calculations undertaken here, the power level is set at 80 percent (20 percent probability of a type II error) and significance level at 95 percent (5 percent probability of a type I error).

³⁸ For example, see Brewer, Crossley & Joyce (2013) or Conley & Taber (2011) for a discussion of the issues surrounding statistical tests for difference-in-difference estimation. The “power” commands in the software package STATA 13 are used here to estimate the MDEs.

³⁹ This proxies the conditions for the estimation of the coefficient on the treatment group in the treatment period when controls are included for time and group dummies in difference-in-difference estimation.

including a smaller number of ineligible three year olds who have only turned three in the term of the survey.

Table 13 presents the impact metric parameters used in the power calculations, derived from the most recently available years of the LFS and FRS. Statistics are presented for the UK for the power calculations for TFC and for England (with a comparison sample of Wales and Scotland) for the FEEE 30 hour extension.

Table 13. Impact metric parameters used in the power calculations

Mean (standard deviation)	Mothers' employment rate (LFS)	Mothers' weekly work hours (LFS)	Fathers' employment rate (LFS)	Fathers' weekly work hours (LFS)	Proportion of children in formal childcare (FRS)
Age group of children					
TFC (UK)					
Aged under 12	65%	27 (12)	88%	41 (10)	32%
Aged 12 or over	76%	30 (12)	82%	40 (12)	n/a
FEEE 30 hour extension (England)					
Aged 1 to 2	57%	27 (12)	89%	41 (10)	49%
Aged 2	56%	27 (12)	89%	41 (10)	58%
Aged 3	56%	26 (11)	89%	41 (10)	72%
Aged 5 to 7	63%	26 (12)	89%	41 (11)	28%
FEEE 30 hour extension (Wales + Scotland)					
Aged 3	65%	27 (12)	90%	41 (12)	72%

Notes: The employment statistics are for parents with children in the age group, but the age 1-2 and age 5-7 excludes parents who also have a child aged 3 and the aged 12 or over group excludes parents with younger children. The LFS data is for 2013 Q4 to 2014 Q3 and the FRS data is for 2013/14.

Table 14 presents the sample sizes used in the calculations for the three surveys. The sample sizes for the analysis using US data are estimated in the following way. HMRC figures indicate that around 1,500 families are estimated to be eligible for TFC in the FRS, which is approximately one third of the 4,500 families with children under the age of 12. Understanding Society data has around 7,000 families under the age of 12, indicating that around 2,300 might be eligible for TFC. Allowing for some change in work and childcare choices which

Annex C: Power calculations

mean some families become eligible for TFC, a scenario of 3,000 eligible families with 5,500 children under the age of 12 is tested in the power calculations. In the power calculations for the RDD approach (using LFS and FRS data), the treatment sample consists of children in the six months after they become eligible for the FEEE 30 hour extension and the comparison group consists of children in the six months before they become eligible. This means that one year of data contains a treatment and control group each equal in size to one half the number of three year old children⁴⁰.

Table 14. Sample sizes used in the power calculations

Age group of children	Mothers (LFS)	Fathers (LFS)	Children (FRS)	Mothers (US)	Fathers (US)	Children (US)
TFC (UK)						
Aged under 12	42,000	32,000		3,000	2,400	5,500
Aged 12 or over	16,000	13,000				
FEEE 30 hour extension (England)						
Aged 1 to 2	8,500	6,500	1,000			
Aged 2	5,000	4,000	500			
Aged 3	5,000	4,000	500			
Aged 5 to 7	11,500	8,500	1,500			
FEEE 30 hour extension (Wales + Scotland)						
Aged 3	650	500	100			

Notes: The estimated sample sizes are based on LFS data is for 2013 Q4 to 2014 Q3 and the FRS data is for 2013/14. The sample sizes for the weekly hours calculations are those for mothers and fathers multiplied by the proportion in work presented in table 13. See text for a description of the derivation of the size of the US samples.

Table 15 presents the estimated MDEs for mothers' work participation and weekly work hours, for fathers' work participation and weekly work hours and

⁴⁰ There is an added complication that a balanced sample across the six months following the eligibility cutoff means that data for the treatment group could not be used from any month prior to 6 months after the introduction of the policy.

for the proportion of children in formal childcare. The primary data option for the matched non-take-up approach for TFC assumes a sample size that is double that of the one estimated for Understanding Society (6,000 families with a child aged under 12 and 11,000 children under the age of 12).

Table 15. Estimated mean detectable effects

Methodology	Mothers' employment rate (percentage point change)	Mothers' weekly hours	Fathers' employment rate (percentage point change)	Fathers' weekly hours	Childcare use (percentage point change)
TFC					
DiD with older children (LFS)	1.1	0.4	1.1	0.4	n/a
Matched non-take-up (actual treatment)					
US:					
- 50% take-up	4.8	1.5	4.2	1.5	3.6
- 90% take-up	9.4	2.3	6.2	2.5	6.0
Primary data:					
- 50% take-up	3.4	1.1	2.5	0.9	2.5
- 90% take-up	5.6	1.8	3.9	1.4	4.2
FEEE 30 hour extension					
DiD with:					
- younger children	2.5	0.8	1.7	0.6	7.7
- older children (LFS + FRS)	2.2	0.7	1.6	0.6	6.7
RDD over:					
- 1 year	3.9	1.1	2.6	0.9	10.5
- 2 years	2.8	0.8	1.9	0.7	7.6
- 3 years (LFS + FRS)	2.2	0.6	1.6	0.5	6.3
Across nations (LFS + FRS)	5.4	1.8	3.5	1.8	12.2

Notes: The figures for mothers' and fathers' employment rate are percentage point changes.

The top half of table 15 considers the MDEs for evaluating the TFC. The considerably smaller MDEs using the LFS reflects the large sample sizes in the survey and the broad age range of the group subject to the policy and in the comparison group. The matched non-take-up approach for TFC has higher MDEs, although it should be considered that these are impacts for actual treatment rather than intention to treat and may therefore be larger. Doubling the

Annex C: Power calculations

sample size through the use of primary data has greater impact on the MDEs is take-up is high (90 percent) and the comparison group smaller.

The lower half of table 15 considers the MDEs for evaluating the FEEE 30 hour extension. The estimates indicate that around three years of the data are required for the RDD approach to generate MDEs of similar magnitude to the DiD approach with younger or older children. The larger MDEs for childcare use reflect the considerably smaller sample sizes in the FRS than in the LFS.

Finally, it should be reiterated that these power calculations are very approximate guidelines and, in particular, do not allow for some of the complexities involved in impact estimations using difference-in-difference approaches. Hence, they should be treated more as indications of the relative merits of the different approaches rather than as firm indications of the size of impacts that can be detected.

References

- Andren, T., (2013), “The choice of paid childcare, welfare, and labor supply of single mothers”, *Labour Economics*, vol. 10, issue 2, April, 133-147
- Averett, S. L., Peters, H. E. and Waldman, D. M., (1997), “Tax Credits, Labor Supply, and Child Care,” *The Review of Economics and Statistics*, vol. 79, no. 1, February, 125-135
- Baker, M., Gruber J. and Milligan, K., (2008), “Universal Child Care, Maternal Labor Supply, and Family Well-Being”, *Journal of Political Economy*, vol. 116, no. 4, August, 709-745
- Bassok, D., Fitzpatrick, M. and Loeb, S., (2014), “Does state preschool crowd-out private provision? The impact of universal preschool on the childcare sector in Oklahoma and Georgia”, *Journal of Urban Economics*, vol. 83, September, 18-33
- Bauernschuster, S. and Schlotter, M., (2015), “Public child care and mothers' labor supply - Evidence from two quasi-experiments”, *Journal of Public Economics*, vol. 123, March, 1-16
- Berlinski, S. and Galiani, S., (2007), “The effect of a large expansion of pre-primary school facilities on preschool attendance and maternal employment”, *Labour Economics*, vol. 14, issue 3, June, 665-680
- Berlinski, S., Galiani, S. and McEwan, P., (2011), “Preschool and Maternal Labor Market Outcomes: Evidence from a Regression Discontinuity Design”, *Economic Development and Cultural Change*, vol. 59, no. 2, January, 313-344
- Bettendorf, L. J. H., Jongen, E. L. W. and Muller, P., (2012), “Childcare subsidies and labour supply: Evidence from a large Dutch reform”, CPB Discussion Paper 217, September
- Blanden, J., Del Bono, E., Hansen, K., McNally, S. and Rabe, B., (2014), *Evaluating a demand-side approach to expanding free preschool education*, Institute for Social and Economic Research.
- Blau, D. M. and Hagy, A., (1998), “The Demand for Quality in Child Care”, *Journal of Political Economy*, vol. 106, no. 1, February, 104-146
- Blau, D. M. and Robins, P. K., (1988), “Child-Care Costs and Family Labor Supply”, *Review of Economics and Statistics*, vol. 70, no. 3, August, 374-381
- Blau, D. M. and Robins, P. K., (1989), “Fertility, Employment and Child-Care Costs”, *Demography*, vol. 26, no. 2, 287-299
- Blau, D. and Tekin, E., (2007), “The Determinants and Consequences of Child Care Subsidies for Single Mothers in the USA”, *Journal of Population Economics*, vol. 20, no. 4, October, 719-741

References

- Blundell, R., Duncan, A., McCrae, J. and Meghir, C., (2000), “The Labour Market Impacts of the Working Families’ Tax Credit”, *Fiscal Studies*, vol. 21, no. 1, 75-104
- Brewer, M., Cattan, S., Crawford, C. and Rabe, B., (2014), *The impact of free, universal pre-school education on maternal labour supply*, Interim Findings, Institute for Fiscal Studies, October
- Brewer, M. and Crawford, C., (2010), “Starting School and Leaving Welfare: The Impact of Public Education on Lone Parents’ Welfare Receipt”, Institute for Fiscal Studies Working Paper 10/19
- Brewer, M., Crossley, T. F., and Joyce, R., (2013), “Inference with Difference-in-Difference Revisited”, IZA Discussion Paper no. 7742, November
- Brewer, M. and Paull, G., (2004), “Reviewing Approaches to Understanding the Link between Childcare Use and Mother’s Employment”, Department for Work and Pensions, Information and Analysis Directorate Research Division, Working paper no. 14, July
- Brewer, M. and Shaw, J., (2004), “Childcare use and mothers’ employment: a review of British Data Sources”, Department for Work and Pensions Working Paper no.16
- Brind, R., McGinigal, S., Lewis, J., Ghezelayagh, S., Ransom, H., Robson, J., Street, C. and Renton, Z., (2014), *Childcare and Early Years Providers Survey 2013*, Department for Education SFR33, September
- Brind, R., Norden, O., McGinigal, S., Oseman, D., Simon, A. and La Valle, I., (2012a), *Childcare and Early Years Providers Survey 2011*, Department for Education Research Report DfE RR240, December
- Brind, R., Norden, O. and Oseman, D., (2012b), *Childcare Provider Finances Survey*, Department for Education Research Report DFE-RR213, May
- Brink, A., Nordblom, K. and Wahlberg, R., (2007), “Maximum fee versus child benefit: a welfare analysis of Swedish child-care fee reform”, *International Tax and Public Finance*, vol. 14, no. 4, May, 457-480
- Bryson, C., Kazimirski, A. and Southwood, H., (2006), *Childcare and Early Years Provision: A Study of Parents’ Use, Views and Experience*, Department for Education and Skills Research Report no. 723
- Cascio, E. U., (2009), “Maternal Labor Supply and the Introduction of Kindergartens into American Public Schools”, *Journal of Human Resources*, vol. 44, no. 1, Winter, 140-170
- Cascio, E. U. and Whitmore-Schanzenbach, D., (2013), *The Impacts of Expanding Access to High-Quality Preschool Education*, Brookings Papers on Economic Activity, Fall.

- Choné, P., le Blanc, D. and Robert-Bobée, I., (2003), “Female Labor Supply and Child Care in France”, CESifo Working Paper No. 1059, September
- Cleveland, G., Gunderson, M. and Hyatt, D., (1996), “Child Care Costs and the Employment Decision of Women: Canadian Evidence”, *The Canadian Journal of Economics*, vol. 29, no. 1, February, 132-151
- Coneus, K., Goeggel, K. and Muehler, G., (2009), “Maternal employment and child care decisions”, *Oxford Economic Papers*, vol. 61, no. S1, April, i172-i188
- Conley, T., G. and Taber, C., R., (2011), “Inference with “Difference in Differences” with a Small Number of Policy Changes”, *The Review of Economics and Statistics*, vol. 93, no. 1, February, 113-125
- Connelly, R., (1992), “The Effect of Childcare Costs on Married Women’s Labor Force Participation”, *Review of Economics and Statistics*, vol. 74, no. 1, February, 83-90
- Davis, E. E. and Li, N., (2005), *Child Care Assistance and the Market for Child Care in Minnesota*, Minnesota Child Care Policy Research Partnership, October
- Davis, E. E., Li, N., Weber, R. and Grobe, D., (2009), *Child Care Subsidies and Child Care Markets: Evidence from Three States*, Oregon Child Care Research Partnership, March
- Daycare Trust, (2008), *Daycare Trust Childcare costs survey 2008*
- Daycare Trust, (2009), *Daycare Trust Childcare costs survey 2009*
- Daycare Trust, (2010), *Daycare Trust Childcare costs survey 2010*
- Daycare Trust, (2011), *Daycare Trust Childcare costs survey 2011*
- Daycare Trust, (2012), *Daycare Trust Childcare costs survey 2012*
- Daycare Trust and Family & Parenting Institute (2013), *Childcare Costs Survey 2013*, February
- Department for Education, (2014), *Statistical First Release: Provision for children under five years of age in England: January 2014*, January
- Department for Work and Pensions, (2014b), *Universal Credit: Increasing the Childcare Offer*, December
- Doiron, D. and Kalb, G., (2005), “Demands for Child Care and Household Labour Supply in Australia,” *The Economic Record*, vol. 81, no. 254, September, 215-236
- Farfan-Portet, M-I, Lorant, V. and Petrella, F., (2011), “Access to Childcare Services: The Role of Demand and Supply-Side Policies”, *Population Research Policy Review*, vol. 30, issue 2, 165-183

References

- Felfe, C., Lechner, M. & Thiermann, P., (2013), “After-school Care and Parents’ Labor Supply”, Centre for Economic Policy Research Discussion Paper no. 9757, November
- Fitzpatrick, M. D., (2010), “Preschoolers Enrolled and Mothers at Work? The Effects of Universal Prekindergarten,” *Journal of Labor Economics*, vol. 28, no. 1, January, 51-85
- Fitzpatrick, M. D., (2012), “Revising Our Thinking About the Relationship Between Maternal Labor Supply and Preschool”, *Journal of Human Resources*, vol. 47, no. 3, Summer, 583-612
- Gelbach, J. B., (2002), “Public Schooling for Young Children and Maternal Labor Supply”, *American Economic Review*, vol. 92, no. 1, 307-322
- Gong, X., Breuing, R. and King, A., (2010), “How Responsive is Female Labour Supply to Child Care Costs: New Australian Estimates”, IZA Discussion Paper no. 5119, August
- Goux, D. and Maurin, E., (2010), “Public school availability for two-year olds and mothers' labour supply”, *Labour Economics*, vol. 17, issue 6, December, 951-962
- Guner, N., Kaygusuz, R. and Ventura, G., (2013), “Childcare Subsidies and Household Labor Supply”, Centre for Economic Policy Research Discussion Paper no. 9775, December
- Gustafsson, S. and Stafford, F., (1992), “Child Care Subsidies and Labor Supply in Sweden”, *The Journal of Human Resources*, vol. 27, no. 1, Winter, 204-230
- Haan, P. and Wrohlich, K., (2011), “Can child care policy encourage employment and fertility?: Evidence from a structural model”, *Labour Economics*, vol. 18, issue 4, August, 498-512
- Hall, S., (2015), *Childcare information for parents*, Department for Education Research Report DFE-RR444, February
- Havnes, T. and Mogstad, M., (2011), “Money for nothing? Universal child care and maternal employment”, *Journal of Public Economics*, vol. 95, December, 1455-1465
- HM Revenue and Customs, (2014), *Childcare Payments Bill 2014 – Impact Assessment*, November
- Huskinson, T., Pye, J., Medien, K., Dobie, S., Ferguson, C. and Gardner, C. with Gilby, N., Littlewood, M. and D’Souza, J., (2013), *Childcare and Early Years Survey of Parents 2011*, Department for Education SFR08/2013, January
- Huskinson, T., Kostadintcheva, K., Greevy, H., Salmon, C., Dobie, S., Medien, K., with Gilby, N., Littlewood, M. and D’Souza, J., (2014), *Childcare and early years survey of parents 2012-2013*, Department for Education SFR06/2014, January

- Jenkins, S. P. and Symons, E. J., (2001), “Child Care Costs and Lone Mothers’ Employment Rates: UK Evidence”, *The Manchester School*, vol. 69, issue 2, March
- Kalb, G. and Lee. W. S., (2008), “Childcare Use And Parents’ Labour Supply In Australia,” *Australian Economic Papers*, vol. 47, no. 3, September, 272-295
- Kazimirski, A., Smith, R., Butt, S., Ireland, E. and Lloyd. E., (2008), *Childcare and Early Years Survey 2007: Parents’ Use, Views and Experiences*, Department for Children Schools and Families Research Report DCSF-RR025
- Kimmel, J., (1998), “Child Care Costs as a Barrier to Employment for Single and Married Mothers”, *Review of Economics and Statistics*, vol. 80, no. 2, May, 287-299
- Kornstad, T. and Thoresen, T. O., (2007), “A Discrete Choice Model for Labor Supply and Childcare”, *Journal of Population Economics*, vol. 20, no. 4, October, 781-803
- La Valle, I., Finch, S., Nove A. and Lewin, C., (2000), *Parents’ Demand for Childcare*, Department for Education and Employment, Research Report RR176, March
- Lefebvre, P. and Merrigan, P., (2008), “Child-Care Policy and the Labor Supply of Mothers with Young Children: A Natural Experiment from Canada”, *Journal of Labor Economics*, vol. 26, no.3, 519-548
- Lefebvre, P., Merrigan, P. and Verstraete, M., (2011), “Public subsidies to private schools do make a difference for achievement in mathematics: Longitudinal evidence from Canada”, *Economics of Education Review*, vol. 30, issue 1, 79-98
- Lundin, D., Mörk, E. and Öckert B., (2008), “How far can reduced childcare prices push female labour supply?”, *Labour Economics*, vol. 15, issue 4, August, 647-659
- Marrufo, G., O’Brien-Strain, M. and Oliver, H., (2003), *Child Care Price Dynamics in California*, Public Policy Institute of California, December
- Michalopoulos, C. and Robins, P. K., (2000), “Employment and Child-Care Choices in Canada and the United States”, *Canadian Journal of Economics*, vol. 33, no. 2, May, 435-470
- Michalopoulos, C. and Robins, P. K., (2002), “Employment and Child-Care Choices of Single-Parent Families in Canada and the United States”, *Journal of Population Economics*, vol. 15, no. 3, August, 465-493
- Müller, K. U. and Wrohlich, K., (2014), *Two Steps Forward - One Step Back? Evaluating Contradicting Child Care Policies in Germany*, SOEP papers on Multidisciplinary Panel Data Research, June
- NDNA, (2015), *NDNA Annual Nursery Survey*, January
- Nollenberger, N. and Rodriguez-Planas, N., (2015), “Child Care, Maternal Employment and Persistence: A Natural Experiment from Spain”, forthcoming in *Labour Economics*

References

- Oishi, A. S., (2002), “The Effect of Childcare Costs on Mothers’ Labor Force Participation”, *Japanese Journal of Social Security*, vol. 1, 51-67
- Paull, G. and Taylor, J. with Duncan, A., (2002), *Mothers’ Employment and Childcare Use in the UK*, Institute for Fiscal Studies, March
- Powell, L. M., (1997), “The Impact of Child Care Costs on the Labour Supply of Married Mothers: Evidence from Canada”, *Canadian Journal of Economics*, vol. 30, no. 3, August, 577-594
- Powell, L. M., (1998), “Part-time Versus Full-time Work and Child Care Costs: Evidence for Married Mothers”, *Applied Economics*, vol. 30, no. 4, October, 503-511
- Powell, L. M., (2002), “Joint Labor Supply and Childcare Choice Decisions of Married Mothers”, *The Journal of Human Resources*, vol. 37, no. 1, Winter, 106-128.
- Rammohan, A. and Whelan, S., (2007), “The Impact Of Childcare Costs On The Full-Time/Part-Time Employment Decisions Of Australian Mothers”, *Australian Economic Papers*, vol. 46, issue 2, June, 152-169
- Ribar, D. C., (1992), “Child Care and the Labor Supply of Married Women: Reduced Form Evidence”, *The Journal of Human Resources*, vol. 27, no. 1, Winter, 134-165
- Ribar, D. C., (1995), “A Structural Model of Child Care and the Labor Supply of Married Women”, *Journal of Labor Economics*, vol. 13, no. 3, July, 558-597
- Rutter, J., (2015), *Childcare Costs Survey 2015*, Family and Childcare Trust
- Rutter, J. and Stocker, K., (2014), *Childcare Costs Survey 2014*, Family and Childcare Trust
- Ryan, R. M., Johnson, A., Rigby, E. and Brooks-Gunn, J., (2011), “The Impact of Child Care Subsidy Use on Child Care Quality”, *Early Childhood Research Quarterly*, vol. 26, no. 3, 320-331
- Schlosser, A., (2011), “Public Preschool and the Labor Supply of Arab Mothers: Evidence from a Natural Experiment”, Tel Aviv University Working Paper, September
- Smith, R., Poole, E., Perry, J., Wollny, I. and Reeves, A. with Cashall, C. and d’Souza, J., (2010), *Childcare and early years survey of parents 2009*, Department for Education Research Report DFE-RR054, October
- Smith, P., Gilby, N., Dobie, S., Hobden, S., Sullivan, L. and Williams, M. with Littlewood, M., D’Souza, J. and Flore G., (2012), *Childcare and early years survey of parents 2010*, Department for Education Research Report DFE-RR221, June
- Viitanen, T. K., (2005), “Cost of Childcare and Female Employment in the UK”, *Labour*, 19 (special issue), 149-179

- Weinraub, M., Shlay A. B., Harmon, M. and Tran, H., (2005), “Subsidizing child care: How child care subsidies affect the child care used by low-income African American families”, *Early Childhood Research Quarterly*, vol. 20, no. 4, 373-392
- Wrohlich, K., (2011), “Labor Supply and Child Care Choices in a Rationed Child Care Market”, DIW Berlin Working Paper no. 1169, November
- Woodland, S., Miller, M., and Tipping, S., (2002), *Repeat Study of Parents’ Demand for Childcare*, Department for Education and Skills Research Report RR348

Frontier Economics Limited in Europe is a member of the Frontier Economics network, which consists of separate companies based in Europe (Brussels, Cologne, London & Madrid) and Australia (Melbourne & Sydney). The companies are independently owned, and legal commitments entered into by any one company do not impose any obligations on other companies in the network. All views expressed in this document are the views of Frontier Economics Limited.

FRONTIER ECONOMICS EUROPE

BRUSSELS | COLOGNE | LONDON | MADRID

Frontier Economics Ltd 71 High Holborn London WC1V 6DA

Tel. +44 (0)20 7031 7000 Fax. +44 (0)20 7031 7001 www.frontier-economics.com