



Feasibility of evaluating the impact of the Access to Work programme

November 2018

Karl Ashworth and Sergio Salis (NatCen Social Research)

Feasibility of evaluating the impact of the Access to Work programme

DWP research report no. 968

A report of research carried out by NatCen Social Research on behalf of the Department for Work and Pensions.

© Crown copyright 2018.

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit <http://www.nationalarchives.gov.uk/doc/open-government-licence/> or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

This document/publication is also available on our website at:
<https://www.gov.uk/government/collections/research-reports>

If you would like to know more about DWP research, please email:
Socialresearch@dwp.gsi.gov.uk

First published 2018.

ISBN 978-1-78659-107-4

Views expressed in this report are not necessarily those of the Department for Work and Pensions or any other Government Department.

Executive summary

Access to Work (AtW) was introduced in 1994 with the aim of supporting people with a disability or long-term health condition to access or remain in work. The programme offers practical and financial support to overcome work-related barriers resulting from this disability or health condition. A number of studies have explored various aspects of AtW but no definitive impact evaluation has taken place. The AtW environment is relatively complex so the Department for Work and Pensions (DWP) has commissioned this feasibility study to undertake an impact evaluation.

The key objectives of the study are to answer the questions:

- Can a robust evaluation of AtW be undertaken with current available data and methodologies; and, if so, how?
- If a robust evaluation cannot be undertaken, what would be required to make it feasible?

The review concludes that the complex challenges facing an evaluation of AtW would require a potentially expensive survey approach to collect the necessary data.

Moreover, especially given the potential cost of the study, further knowledge should first be enhanced on a number of issues in order to help better design the evaluation:

- Understand better the triggers and trajectories of the recipients journey into an AtW claim and approval
- Get a better understanding of employers' use of reasonable adjustments and how this interacts with AtW claims and approvals
- Understand better the caseworker decision making to help inform how to classify people into the comparison group from the pool of potentially eligible AtW non-recipients
- Consider pilot work for a survey exploring interview modes and operationalising questions to detect triggers and trajectories into AtW
- Produce a robust estimate of the size of the population of workers with health conditions who meet the AtW eligibility conditions but do not claim AtW

The key challenge for any impact evaluation is estimating 'business as usual', i.e. the counterfactual of what would have happened in the absence of treatment. Various approaches to create the counterfactual are considered and statistical matching was deemed the approach best suited to an evaluation. Methods requiring pre-intervention measures were ruled out because of the long-standing nature of AtW: pre 1994 conditions were considered unsuitable for creating a counterfactual for AtW in the present day. No obvious instrument was identified for an instrumental variables approach and regression discontinuity was deemed inappropriate because eligibility for AtW does not meet the necessary criteria enabling a cut-off point on a measure for allocation purposes.

Feasibility of evaluating the impact of the Access to Work programme

Matching requires the existence of a group of people who meet the eligibility conditions for AtW but who have not applied. Given that in 2017 around 3.5 million people in work reported a disability, it seems likely that a substantial number would meet the eligibility conditions for AtW. In many cases, it is assumed that changing health conditions will be a trigger for applying for AtW, though it is also anticipated that changes in other circumstances may also be influential for some workers already facing long-term health problems affecting their ability to work. What will be beneficial before committing to an evaluation of AtW is improved knowledge of triggering events, the extent to which triggers lead to AtW claims and reasons why they do not. Similarly, it is important to establish the degree of awareness of AtW among workers with health issues.

Further complicating the matching process is the role of reasonable adjustments which employers have a duty to implement under the 2010 Equality Act. Not only are reasonable adjustments inherently subjective in determining if a duty arises for the employer, smaller employers can reclaim more of the costs of implementing AtW than larger employers are able to. These complexities arising from employers and their reasonable adjustment duties and variable AtW compensation make it even more difficult to identify appropriate comparison cases for matching to AtW recipients.

An AtW evaluation would require longitudinal data. In addition to identifying the relevant matching data, which for AtW applicants must be collected prior to the approval of AtW, data are also required to measure outcomes over a period subsequent to AtW treatment.

An exploration of existing administrative and survey data showed that the current available data were of an insufficient sample size and typically failed to collect the full range of matching data required.

An evaluation of AtW would require a bespoke survey to collect the relevant data. Such a survey would be complex in that it would require a dual frame, selecting AtW recipients from DWP AtW records, and a separate frame for a comparison group of workers. The comparison group could initially be identified as workers from HM Revenue and Customs tax records. However, a potentially large-scale screening exercise would be required to identify those people with appropriate health challenges to filter through to the comparison group. Such an exercise would likely prove costly and may require administering a web-mode survey, though assisted technologies would be required for visually impaired workers. Initial exploratory work would also be required to understand better quality issues, such as non-response and mode bias. Consequently, thorough piloting of design issues is recommended prior to undertaking a potentially expensive evaluation.

Contents

- Executive summary 3
- Contents 5
- Acknowledgements 7
- 1 Overview 8
- 2 Introduction..... 9
 - 2.1 Estimating business as usual 9
 - 2.2 Other impact assessment challenges..... 10
 - 2.3 Structure of report 11
- 3 Background: The Access to Work programme 12
- 4 Key evaluation aspects and concepts 15
 - 4.1 Eligibility for Access to Work 15
 - 4.1.1 Eligibility and award rules..... 15
 - 4.1.2 Identifying non-recipients of AtW..... 16
 - 4.2 Outcomes of interest 18
 - 4.3 Access to Work recipient typologies and heterogeneity 21
- 5 Methodological options..... 23
 - 5.1 Defining treatment and the counterfactual..... 23
 - 5.2 Treatment effect of interest and impact estimation methods 24
 - 5.2.1 The application process 25
 - 5.2.2 The fulfilment of the eligibility conditions and receipt of treatment 25
 - 5.2.3 Statistical matching and challenges 27
 - 5.3 Measuring the counterfactual 27
 - 5.4 Observation window for the study 29
- 6 Data options 30
 - 6.1 Secondary survey data..... 30
 - 6.2 Administrative data..... 32
 - 6.3 Primary data collection 33
 - 6.3.1 Calculating sample size 33
 - 6.3.2 Sample design 34

Feasibility of evaluating the impact of the Access to Work programme

- 6.3.3 Data collection mode..... 35
- 7 Exploiting changes in Access to Work conditions..... 37
 - 7.1 The introduction of Personal Independence Payments..... 37
 - 7.2 Selective marketing of Access to Work 38
- 8 Gaps and further research..... 39
- 9 Appendix: Bibliography of Previous Access to Work Research..... 41
- 10 References 42

Acknowledgements

The project team would like to thank all the participants in the Access to Work feasibility study workshop, which took place in London on the 26th of March 2018. These were Mike Daly, James Halse, Gordon Pal, Mark Walsh, David Irvine and Imogen Butcher (Department for Work and Pensions), Paul Bivand (Learning and Work Institute) and Stephen Morris (Manchester Metropolitan Institute). We also owe a debt of gratitude to Erica Allaby for helping us to understand better the caseworker process in their evidence gathering and approval of claims for Access to Work.

Any errors and omissions remain the responsibility of the authors.

Views expressed in this report are not necessarily those of the Department for Work and Pensions or any other government department.

1 Overview

Access to Work (AtW) was introduced in 1994 with the aim of supporting people with a disability or long-term health condition to access or remain in work. The programme offers practical and financial support to overcome work-related barriers resulting from this disability or health condition. Being in work has obvious advantages to people themselves (e.g. improved confidence, income, and mental health and wellbeing), and if the benefits of the AtW programme exceed its costs the whole economy will benefit from a financial boost. Therefore, from a policy perspective, evaluating the impact of the programme on the individuals covered by its provision is of paramount importance.

The Department for Work and Pensions (DWP) commissioned NatCen Social Research to explore the feasibility of carrying out a robust evaluation of the impact of AtW, and the primary objective of this study is reporting the outcome of this assessment. This feasibility study benefited from a workshop which brought together NatCen's research team, DWP staff and other work and evaluation experts.

Previous research has been undertaken on AtW; a list is given in Appendix A. Notably, the Sayce¹ review reported that AtW recovered £1.48 for every £1.00 spent but it is not clear how robust this is as an estimate. No rigorous impact estimation has been previously undertaken, which is understandable given the challenges outlined in this report.

The key objectives of the feasibility study are to answer the questions:

- Can a robust evaluation of AtW be undertaken with current available data and methodologies; and, if so, how?
- If a robust evaluation cannot be undertaken, what would be required to make it feasible?

¹ Sayce (2011) *Getting in, staying in and getting on DWP*

2 Introduction

It is appropriate first to consider the necessary conditions required to conduct an impact evaluation. By which we mean a study aimed at detecting whether or not a policy intervention (the ‘treatment’) has had an effect on intended outcomes; and if so, determining the direction and magnitude of such an effect. It is also important to consider the potential for the impact to have differential effects, i.e. a larger or smaller magnitude or opposite directions, for different groups of people.

2.1 Estimating business as usual

The key challenge for any impact evaluation is estimating ‘business as usual’, i.e. the counterfactual of what would have happened in the absence of treatment (i.e. without Access to Work (AtW)). The ‘fundamental problem of causal inference’ (Holland, 1986) arises from the fact that we cannot simultaneously observe the same unit in two different states, i.e. with and without treatment, to measure directly the impact of the treatment. Rather, the counterfactual is estimated using the most appropriate data and methods possible with the available resources and operational constraints.

In principle, the counterfactual may be estimated using a comparison group, alongside the treatment group, through collecting outcome data for both groups at the same time period subsequent to the implementation of the treatment. This is the basic two-group, post-intervention design for randomised control trials and statistical matching techniques. Regression discontinuity and instrumental variable designs also typically operate using this approach. Another way of forming the comparison group is to use outcome observations collected prior to the introduction of the treatment, which are then compared to measures of the outcome subsequent to the treatment. Interrupted time series designs use this approach but are susceptible to confounding through coincidental changes in other causal factors occurring temporally alongside the treatment intervention. Difference-in-difference (DiD) techniques use both pre and post intervention time points with a treatment and comparison group to help subtract out potential systematic confounding change arising from changes in other causal factors.

The evaluation of a long-standing and well-established policy, such as AtW, constrains the available methodological options compared to what is possible when evaluating a new policy; the impact of which can, in principle, be tested before rolling out nationally. Starting the evaluation when the policy is well established severely limits what can be done using pre-treatment time periods to help estimate the counterfactual. It may be possible to assess the impact of naturally occurring changes to AtW within the timeframe covered since policy inception. However, the nature of the counterfactual and the meaning of the subsequent impact effect will be different from that which would be estimated from a baseline of no policy intervention.

In undertaking this feasibility study, we have focused primarily on assessing methods which would provide a counterfactual based on a no policy intervention because this would give the most authoritative impact estimate for the effect of AtW. In general, we believe that the most promising approach to estimate a counterfactual is through statistical matching techniques. Methods using pre-intervention time periods are not sensible given the long duration of AtW's existence; and no instrument has been identified to make use of an instrumental variables approach for a counterfactual scenario where AtW does not exist. Even so, statistical matching is still challenging, as is discussed below. However, we have also sought to outline evaluation opportunities arising from natural changes affecting subgroups of the AtW population. In such cases, the counterfactual is clarified at the point of discussion.

It is important to keep in mind that the 'business as usual' scenario is quite complex. We discuss below 'reasonable adjustments', which employers are required by law to make to help people with health and disability challenges to work. Other benefits, such as Disability Living Allowance (DLA) and Personal Independence Payment (PIP) may also be available to some people in the target population. Consequently, we need to be alert to how AtW may interact with other policies and benefits when considering both treatment effects and creating comparison groups.

2.2 Other impact assessment challenges

Assessing the potential to implement a robust impact assessment of AtW entails a number of evaluation challenges. These include, for example, identifying a population of recipients for which the programme's impact can and needs to be estimated (e.g. those eligible for or receiving AtW support), defining successful outcomes achievable through AtW provision and exploring whether existing data can provide operational measures of such outcomes. These are in addition to the challenge of identifying an appropriate counterfactual outcome to provide a baseline against which the success of the programme can be measured. This crucially relies on finding a comparator group of untreated individuals (i.e. non-AtW recipients) who resemble the treated group across a number of relevant characteristics (notably, the eligibility conditions).

In setting out the challenges that the impact analysis of AtW is most likely to encounter, this feasibility study outlines data options available to the evaluators. The data to be used for an impact assessment must fulfil different roles, such as:

- Identifying treated individuals (broadly speaking, AtW recipients)
- Providing appropriate variables to select a comparison group of untreated subjects (non-AtW recipients)
- Measuring relevant outcomes for both AtW and non-AtW recipients

In order to meet these conditions, candidate data sources will have to be comprehensive in terms of coverage of subject domain variables and have a large sample size to measure a meaningful impact between treated and untreated individuals. These are relatively demanding conditions, which limit the available data options. This report looks at existing secondary data sources and reviews the extent

to which they meet these conditions and considers what would be required to run a primary survey data collection exercise.

2.3 Structure of report

This report is organised as follows. Section 2 provides a brief introduction to the AtW programme, and illustrates recent trends based on available statistics from the Department of Work and Pensions (DWP). Section 3 outlines the key evaluation aspects of the programme and other theoretical concepts that need further exploration to inform the choice of the most appropriate impact estimation design. Sections 4 and 5 explore available methodology and data options, respectively. Section 6 explores the evaluation potential of naturally occurring changes in circumstances for people eligible for AtW. Finally, Section 7 offers some suggestions for future research to help fill current knowledge gaps.

3 Background: The Access to Work programme

The Access to Work (AtW) programme was introduced in Great Britain in 1994 with the aim of supporting people with disabilities or long-term health conditions to start or continue in work. The programme seeks to remove work-related barriers, ensuring that disabled individuals are not in a position of disadvantage in the workplace compared to other employees who are not disabled. AtW works alongside the reasonable adjustments, which employers are required to make as a consequence initially of the Disability Discrimination Act, 1995; and reinforced by the Equality Act, 2010.

Reasonable adjustment is somewhat subjective as a term and is explicitly aimed to reflect individual circumstances. The Equality Act 2010² states:

‘In the case of employers, whether a duty arises will depend on the circumstances of each individual case. There is no duty owed to disabled people in general in an employment context.’

AtW is intended to supplement reasonable adjustments and is not a substitute payment source. However, an employer’s legal duty to make a reasonable adjustment takes into account their resources to fund the change. Consequently, it is possible that employees working for low-resourced employers may get AtW to cover adjustments that better resourced employers would be expected to fund themselves. Moreover, it is likely that employers will also vary in their willingness to make reasonable adjustments, irrespective of their legal duty. Potentially, this ambiguity and the subjectivity surrounding reasonable adjustments may have consequences for defining the counterfactual for AtW, which are discussed further below.

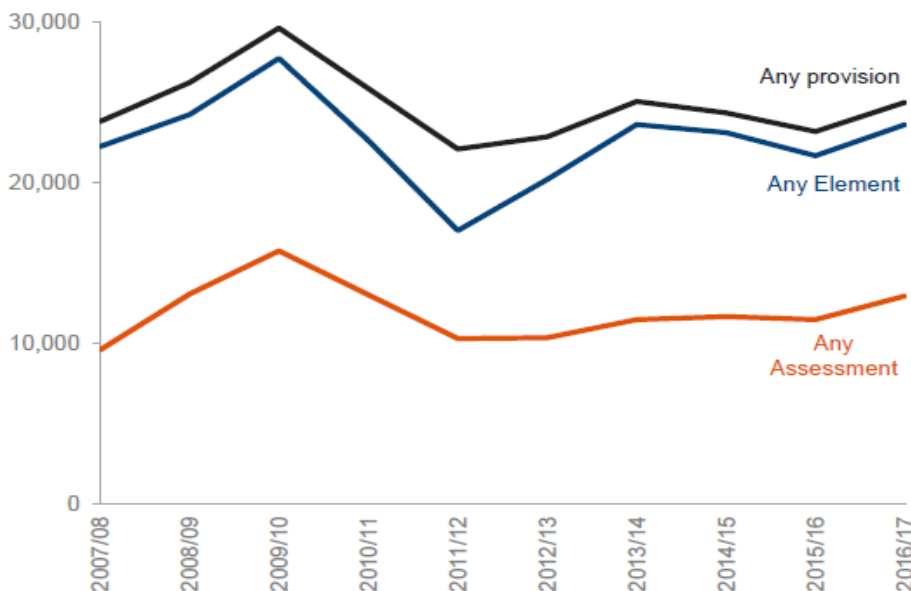
Upon application for AtW, the DWP either rejects or approves the application after considering the applicant’s eligibility and the employer’s duty to make reasonable adjustments. Upon approval, the applicant may receive one or both of two types of AtW provision: assessments and elements. Assessments aim to explore workplace-related barriers and make recommendations on possible ways to overcome them. Elements seek to supplement the reasonable adjustment made by employers in a number of different ways, such as through provision of communication support for interviews, payment of travel costs, special aids and equipment, adaptations to premises and vehicles, and access to mental health services.

² [The Equalities Act 2010](#)

Feasibility of evaluating the impact of the Access to Work programme

Recent data show that most (94% per cent) of those with AtW provision approved in 2016/17 (25,020 individuals) were recorded under the elements strand (see Figure 1), and assessments were in some cases a pre-requirement for element approval.

Figure 1 Number of recipients with AtW provision approved



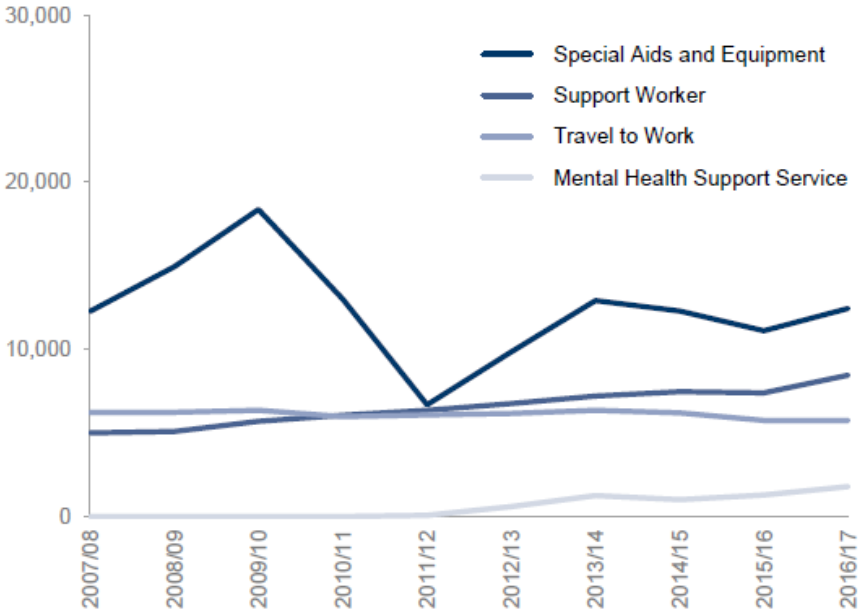
Source: Access to Work statistics (DWP)³

Note: provision refers to a new claim in the given financial year (this may be a new, renewal or repeat recipients). Recipients may be included more than once in any year, e.g. Any Provision and Any Assessment, Any Provision and Any Element, Any Provision and Any Assessment and Any Element.

The general trend in the number of recipients who had one or more elements approved is largely driven by one particular element, namely special aids and equipment (SAE), see Figure 2. In the financial year 2016/17, around 43 per cent of those with any elements approved were supported through this type of provision; around three in ten had a support worker and one-fifth had travel to work costs paid. The trend for SAE approvals is volatile, peaking in 2009/10, reaching its lowest in 2010/11; after which it rises again to around the same number seen in 2007/08. The number of those who had a Mental Health Support Service (MHSS) element approved has gradually been increasing since this service was introduced in December 2011, reaching 7.5 per cent in the financial year 2016/17. Approval of a support worker is on a slow upward trend and has gradually been increasing since 2007/08. In contrast, support for Travel to Work has declined slightly since 2007/08. Approvals of other support types such as adaptations to premises/vehicles or travel in work were negligible in comparison to the four main elements.

³ [DWP Access to Work Statistics](#)

Figure 2 Number of recipients with the most frequent approved AtW element types



Source: Access to Work statistics (DWP)

Interestingly, as Figure 2 shows, the proportion of AtW recipients who received support through the provision of special aids and equipment experienced a big dip in the financial year 2011/12. The reason for this sharp fall is not known with certainty, but the AtW statistical release suggests the dip may be due to revised guidance clarifying the distinction between provision that can be approved by AtW and that made through reasonable adjustments. Whatever the reason, we propose that in order to avoid distortion from any impact of the Equality Act on AtW behaviour, the impact analysis should be restricted to the more recent period (i.e. post-2011/12).

4 Key evaluation aspects and concepts

4.1 Eligibility for Access to Work

An important step in any impact evaluation is to define the population of recipients for which estimating an impact is relevant from a policy perspective. Broadly speaking, here this means identifying the group of people who meet the eligibility conditions for Access to Work (AtW). This then allows the identification of the treated group, who are AtW recipients⁴ and an untreated group, i.e. eligible non-recipients. Currently, the size of the AtW target population is unknown, largely because the size of the eligible non-recipient group is unknown. Establishing these unknowns is important prior to committing to an evaluation study, because the potential to create a comparison group from which to estimate a counterfactual depends upon the availability of sufficient numbers of eligible non-recipients.

Establishing AtW eligibility conditions is therefore crucial in order to identify the AtW eligible population, i.e. recipient and non-recipient AtW recipients. AtW recipients can be identified directly from their claim status from Department for Work and Pensions (DWP) administrative records. However, it is equally important to ensure that any individuals considered for inclusion into the pool of potential comparators are appropriately selected, albeit from other sources. To achieve equivalence between the AtW recipients and the comparison group we must understand the eligibility conditions for AtW in order to replicate them for use in identifying the eligible AtW non-recipients in the context of available data sources.

4.1.1 Eligibility and award rules

AtW has a complex set of eligibility conditions and awards rules. A full list of eligibility conditions is given in the Access to Work: staff guide online document⁵. However, the main criteria to be granted AtW include the following eligibility conditions:

- Having a disability or (physical and/or mental) health condition that limits their ability to work. AtW guidance states that the limiting condition should be expected to last a minimum of 12 months (long-term health condition)
- Being resident in Great Britain (England, Scotland or Wales)
- Being aged 16 or over⁶

⁴ A more detailed discussion of the definition of AtW recipients and how this relates to treatment effects is undertaken in the following section.

⁵ [Access to Work: staff guide](#)

⁶ While the eligibility criteria do not prescribe any upper limit on individuals' age, outcomes will have to be observable for a certain period following treatment and therefore any impact assessment will have to exclude the oldest employees.

Feasibility of evaluating the impact of the Access to Work programme

- Be in a paid job (this includes paid employment, self-employment, apprenticeship, work trial/experience and internship) or about to start one
- Be earning at least the National Living Wage or National Minimum Wage rate for each hour worked
- Under Jobseekers Allowance (JSA), Universal Credit or Income Support, a claimant must work more than one hour a week
- Under Employment Support Allowance it is necessary to be doing permitted work of less than 16 hours per week, earning up to £125.50 per week, which has been agreed with the work coach

There is a complex relationship between receipt of welfare benefits, work and eligibility for AtW; as some restrictions apply if the individual is getting certain benefits. For example, those who are already claiming Universal Credit, Jobseeker's Allowance or Income Support can get AtW support only if they work more than one hour a week. Employment Support Allowance claimants can get AtW support only if they are doing permitted work, which implies earning not more than £125.50 a week and working less than 16 hours a week, and AtW support is agreed with their work coach. AtW is also available to people receiving Personal Independence Payment or Disability Living Allowance.

The AtW caseworker first has to determine the applicant's eligibility for AtW, taking into account the eligibility criteria, and then to determine whether or not a reasonable adjustment is sufficient to meet needs. Applicants meeting the eligibility criteria and whose needs exceed reasonable adjustments are the target population of interest for the evaluation; informally, it is those workers whose health condition puts them at risk of losing their employment.

4.1.2 Identifying non-recipients of AtW

The imperative to identify eligible non-recipients of AtW means being able to replicate the objective eligibility conditions for AtW. It also requires data to emulate the AtW caseworker's judgement regarding the health condition and judging the appropriateness of reasonable adjustment as a resolution. In principle, the objective conditions can be identified through a survey questionnaire; or may be captured in administrative data sources. More challenging is identifying and collecting appropriate data to permit emulating the AtW caseworker's judgement of the person's health condition and the potential for reasonable adjustment as a resolution.

In evidence gathering for consideration of an AtW claim, a caseworker will discuss an applicant's needs with them, determining what support is required. This will require an independent health assessment for travel to work, but a health assessment is at the discretion of the AtW caseworker otherwise. Caseworkers may contact employers directly to suggest reasonable adjustments but there may also be a need for a workplace assessment with the employer, to help identify the most suitable type of provision (if any). Larger employers are required to share the cost of the AtW provision (up to a maximum amount), but smaller companies are not. Crown employers are now expected to cover AtW costs in full.

Feasibility of evaluating the impact of the Access to Work programme

It is clear from the above discussion that there are implications for identifying eligible non-recipients of AtW arising from AtW's complex relationships with work and benefits, and the subjective aspects of identifying health conditions appropriate to AtW and determining which elements are appropriate. In principle, the relationship between AtW eligible work and benefits conditions can mostly be captured because, though complex, these are objective conditions which can be assessed within a survey context; though data requirements are detailed and data quality needs to be high. More challenging is identifying, and filtering out from the comparison group, workers with health conditions where AtW support is not considered necessary, or those whose needs may be resolved through encouraging employers to make reasonable adjustments. From the perspective of non-applicants, such events are hypothetical and identifying predictive indicators within a survey context to estimate the likely outcomes, should a person apply for AtW, is challenging.

What is less clear is the extent to which AtW applicants are rejected in their AtW applications and why. Similarly, we do not know how many employers implement reasonable adjustments as a consequence of contact with an AtW caseworker or an assessment arising from an AtW claim. Within a statistical matching paradigm these unknowns have implications for matching purposes, estimating the counterfactual and defining the treatment effect; which are discussed in more detail below. However, the extent of their influence will depend in part upon the prevalence of such occurrences. If few people are affected by these events then their influence will be small. Conversely, a larger prevalence implies a larger effect. For these reasons, it would be helpful to find out more about the application process, the number of rejected applications and the reasons for rejection, and employer actions as a result of an assessment (particularly in the absence of an element being awarded). Qualitative research exploring the process for caseworkers may help inform these issues.

Similar considerations apply to employer's support. Ideally, any help received in the workplace by both AtW recipients and their untreated counterparts should be accounted for in the impact analysis. In principle, it is also possible that an employer is willing to make adjustments beyond what they considered 'reasonable' without seeking AtW funding or, conversely, be unwilling to make such adjustments when expenses are covered through AtW funding. However, capturing these circumstances will be extremely difficult in practice. It is plausible that differences in employers' willingness and/or capability to provide support (regardless of whether this qualifies as reasonable adjustments) will be influential in determining whether eligible employees make an AtW claim, and specific employer's features may allow the evaluator to control for such differences. For example, large or public firms may be more likely to make adjustments and/or provide support to those with a disability or health condition than small and private businesses; likewise businesses with a Human Resources function. Reasonable adjustment may also be more likely to be made for those in strategically important occupations (e.g. managerial roles).

It will therefore be important to understand how the selection process is affected by the subjective assessment by caseworkers and what features are readily observable (different impact estimation techniques may be required if selection is based on

unobservable factors). At this stage, we assume that most of the eligibility criteria are readily observable and those which cannot directly be observed by the evaluator can be proxied by using existing information (for example, the propensity to make reasonable adjustment and other support could be reflected in the employee's occupation and/or their employer's size and sector). This justifies a preference for a matching approach over other more complex alternative estimation methods (see the following section on 'Methodological options'). It is also important to note that what is observable in the context of matching is primarily dependent upon data availability. Bespoke surveys can, in principle, be designed to collect any data. However, some concepts are more difficult to capture with high quality than others, particularly where recall of events is involved and recall bias and misremembering can occur.

4.2 Outcomes of interest

The individual outcomes that are most likely to be used in an AtW impact evaluation are the following:

- Employment⁷ retention (conceptually defined in relation to an individual continuing to work for the same employer or changing employer, with no substantive break between employers⁸)
- Employment advancement (e.g. the AtW provision helps the individual to get a promotion or a better job)
- Improved health (i.e. the individual's physical and/or mental health improves as a result of the AtW)

Job retention can be considered in a number of ways, including time spent in a job spell with a single employer (or as self-employed), time aggregated across multiple contiguous work spells (also across different employers), or a simple aggregation of total days spent in work as a proportion of total time available for work, within a given time period. For the purposes of the present study, there appears to be little gain from distinguishing between continuous work spells with a single employer from contiguous work spells with multiple employers⁹. It will be important to account for any unemployment/inactivity gaps between two contiguous work spells, either explicitly (by defining the maximum acceptable gap) or implicitly (a continuous

⁷ Here, 'employment' is used generally to include working for an employer and/or working as self-employed. Moreover, it is also possible that people on AtW may hold more than one job at the same time.

⁸ It may be important to explore the extent to which a change in employer correlates with a new application for AtW, perhaps to meet different working conditions arising from new employment circumstances.

⁹ However, we do note that an AtW recipient may put in a new claim for a new employer, which might be classified as a 'repeat' claim.

Feasibility of evaluating the impact of the Access to Work programme

measure of employment retention may reflect the gap length as longer gaps will result in lower proportions of days spent in work over a given period).

The appeal of a work spell-based outcome measure is that it enables the use of event history analysis techniques, for example survival analysis. However, the approach of taking the aggregate proportion of days worked in a given time period has the potential advantage of simplicity, and we can use statistical tests appropriate for differences between means, which may be more easily understood by non-specialist audiences.

With respect to employment advancement, defining a 'better' job may be too subjective and ambiguous an exercise. Advancement can be defined as a promotion, which potentially could be measured through increased average hourly earnings (the increase could be observed between two contiguous work spells or between/within pre-defined follow-up periods). Some of the benefits attached to a 'better' job and which are hard to measure (e.g. lower job-related stress, positive working environment and reduced travel to work time) would be expected to lead to greater employment retention. Therefore, in the longer term the impact of the AtW on employment retention may indirectly capture some of these benefits.

Outcomes denoting health improvements may also be difficult to measure beyond general health and wellbeing. In principle, AtW recipients are expected to experience improved satisfaction as a result of the support received through AtW, and this should be reflected in any self-assessed health measure. General Health Questionnaires¹⁰, subjective health measures based on Likert scales or other indices of satisfaction in the workplace could be used to measure individuals' wellbeing and mental health. The most obvious outcome to be used in the impact assessment would be the proportion of AtW recipients screened as having positive (or negative) wellbeing or mental health, but continuous measures of mental health could also be constructed. There could also be scope for physical improvement for recipients with certain physical conditions, but it is anticipated that this effect of the AtW will concern only a limited subpopulation of recipients. However, from the perspective of planning an evaluation of AtW's effect on health outcomes, it is immediately clear that survey data would be required because administrative data do not contain the measures discussed above.

Another potential outcome of interest is the role of AtW in supporting job starts but given that this would require the construction of a different counterfactual we do not consider this option here. Moreover, AtW largely supports people already in work; and, consequently, we suggest that any impact evaluation will gain more from focusing on retention and advancement.

The relationship between health, employment and AtW is likely to be complex and a-priori it is not clear to what extent we can assign a causal relationship between AtW and health improvements. Nevertheless, we anticipate that it will be important to establish basic transition probabilities and sample sizes, as well as establishing event trajectories linking health, employment and AtW receipt over successive time points.

¹⁰ [General Health Questionnaires](#)

Feasibility of evaluating the impact of the Access to Work programme

4.3 Access to Work recipient typologies and heterogeneity

AtW recipients are a heterogeneous group of people, and any impact analysis will have to capture individuals' heterogeneity in the most appropriate way, such as controlling for individual characteristics which result in a higher likelihood of making a successful AtW application, or exploring the impact of the programme separately for different recipient segments or types by means of subgroup impact analyses. The key issue we consider here is the extent to which we can match within groups, such as health condition, employment size, age and gender. The alternative to direct matching is to enter matching variables into a distance metric such as a propensity score and to match using that propensity score. Most likely, the matching will use a small number of stratification variables for grouping purposes and then use propensity score matching within each of these groups. Identifying which variables are appropriate for stratification is the aim of this sub-section.

As discussed previously, AtW recipients differ across the types of support received, the most common types being the provision of special aids and equipment, support workers, help with travel to work, and access to mental health support services. Reflecting the support received, AtW recipients differ across the health conditions that made them eligible for treatment. Available DWP statistics indicate that, among those who had AtW support approved in 2016/17 (either assessments or elements), the most reported health conditions were deaf/hard of hearing, back or neck, Dyslexia and difficulty in seeing. Mental health conditions¹¹ were less frequently reported compared to physical conditions (under-reporting may, in part, be due to the stigma attached to such conditions). Challenges in attempting to classify people with health conditions into broad typologies are discussed below. Given these challenges, entering health conditions and history variables into the propensity score calculation appears a more productive approach than using broad health typology groupings to stratify the matching process.

Age and gender subgroups may also be of interest. Among those who had AtW support approved in 2016/17 (either assessments or elements), the majority (60 per cent) were aged 40 or over, and 60 per cent were women.

The link between employer size and AtW liability for contributing to AtW awards was discussed earlier and for this reason, it may be sensible to at least explore employment size and sector as potential stratification variables for matching.

¹¹ AtW only records the primary health condition, so it is possible that the prevalence of AtW recipients with mental health issues is greater than is suggested by the AtW data because people may have both physical and mental health issues.

Feasibility of evaluating the impact of the Access to Work programme

AtW recipients relate to employment in a number of different ways as the programme includes those who are about to start a job, people requiring support for an internship, Work Trial or job interview, and those who experience difficulties in their existing job. Employment can be either paid or self-employment. As previously explained, this feasibility study will mainly focus on AtW recipients already in work. While the majority of individuals are expected to be paid employees, a comparison between paid and self-employed (sample sizes permitting) may be of interest.

Another way to classify individuals supported by AtW is by referring to their claiming status at a particular point in time or within a pre-specified period of observation. For example, within a given time window, AtW recipients can be considered new recipients if they start receiving AtW support for the first time, repeat recipients if they start being supported having already received AtW support in the past or existing (renewal) recipients if they are already receiving help through AtW before the beginning of the chosen time window. Depending on the time window chosen, stock (existing) and flow (repeat and new) recipients may be affected by the introduction of the cap.¹² However, research conducted by DWP has shown that this cap has affected only a very small proportion of recipients (around one per cent, some 200 people in total), and therefore its introduction is not expected to pose a threat to impact analysis.

¹² On the 1st of October 2015 a cap (ceiling) of £40.8k per annum (1.5 x average earnings) was introduced (this increased to £57.2k – twice average earnings from April 2018). A possible implication of the cap (should this have a negative effect on some recipients) could be that the average impact of the AtW in the period before the 1st October 2015 was larger than the impact observed afterwards but this may not be the case if the protection arrangements in place for some people offset the negative effects of the cap for the individuals concerned. Possible anticipation effects should also be considered as the cap was announced in advance and this might have discouraged some individuals from seeking support and/or prompted others to make a claim.

5 Methodological options

5.1 Defining treatment and the counterfactual

It is outside the remit of this report to develop a detailed theory of change for Access to Work (AtW) but it is still appropriate to consider broadly how AtW might have an impact. From the above discussion on the heterogeneity of recipients, AtW offers work-enabling support in a variety of ways encompassing AtW recipients facing different challenges. It is apparent that the nature of the health condition and size of the employer are key influences on how AtW will work. However, whilst we can group employers following the categories used by AtW to define legal responsibilities for sharing AtW costs, grouping people appropriately by their health conditions is far more challenging. There are two key issues in defining an appropriate health categorisation. The first challenge relates to the need to create a categorisation that is independent of the type of AtW received, i.e. it can be applied to both the treatment and comparison group. Consequently, it will not be possible to use AtW process data in this context. The second challenge arises from inherent difficulties arising from the potential multiplicity of co-occurring health conditions. Increasing the number of health categories to capture more accurately the diversity of health conditions will tend to result in smaller sub-populations and require ever larger sample sizes to undertake a robust evaluation.

From this perspective, we can consider treatment to coincide with the receipt of AtW; i.e. the approval of a claim. As outlined above, we can distinguish claim approvals between three applicant types:

- New: have never received an AtW award previously
- Repeat: have previously received AtW, but have had a period of non-receipt between this and the end of the previous award
- Renewal: the current award is made immediately after the expiration of the previous award

Ideally, these applicant types would be considered separately for evaluation purposes because the impact effect could vary by type. However, sample size limitations might require pooling them into a single group for analytic purposes. Under such conditions, arguably, it would be preferable to remove renewal recipients, if possible, because the effect for long-term recipients could be very different from that for new recipients.

Attempting to define treatment raises a number of questions. One question is whether or not to include action taken by an AtW caseworker to contact an employer and resolve the applicant's need through reasonable adjustment made by an employer, without the need for an AtW award. Arguably the adjustment would not have been made without intervention from the AtW caseworker, so could be

classified as an impact. However, such outcomes are only recorded in caseworker management information data and would have to be recovered from that source.

It would also be possible to treat elements as distinct from an assessment (with no element awarded). The assessment award can indicate various outcomes including payment by an employer, coverage by a government department and situations where a need is present but no support can be provided. In principle, it would also be possible to treat separate elements as different types of treatment, operating in different ways to define separate impact effects.

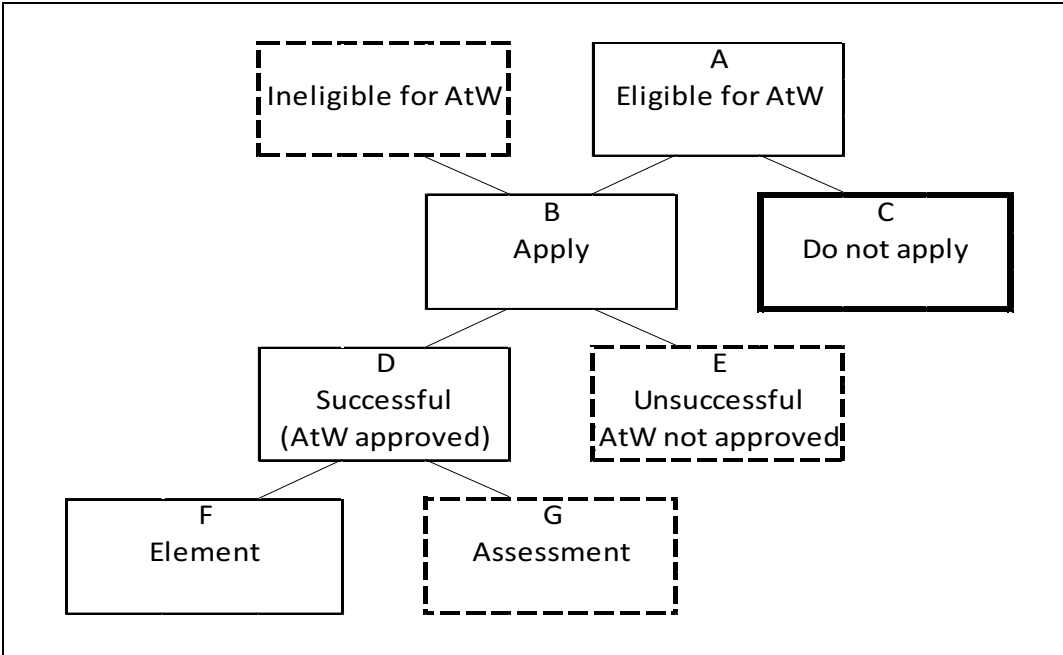
In addition, there are practical issues in trying to create a counterfactual for each of the different population sub-groups. We would expect these different population subgroups to share a number of similarities and also to have some distinguishing characteristics setting them apart from other subgroups. In order to get a good match we would need to understand, and have data for, each of these sets of distinguishing characteristics to get an appropriate match when creating the counterfactual for each subgroup. Further analysis of AtW recipients would be required to establish how well we can distinguish between different recipient subpopulations using available statistical data on AtW recipient characteristics. Such an analysis would give a better idea of the practicality of making such fine distinctions in the evaluation. However, it would also be necessary to ensure a large sample size for each recipient subgroup, and their comparison counterpart group, to enable a robust estimate of the impact effects in the evaluation.

A standard approach in impact evaluation designs when dealing with complexity arising from many 'moving parts', such as demonstrated here, is to estimate an average treatment effect across the population group of interest rather than trying to unpack quantitatively the 'black box' of what works for whom and how. This standard 'black box' treatment is considered here to be the most promising approach given the challenges in dealing robustly with the heterogeneity of recipients and the potential multiplicity of ways in which AtW could work.

5.2 Treatment effect of interest and impact estimation methods

Different methodologies can be used to estimate the AtW impact, each one providing the estimate of a different treatment effect (impact) parameter which concerns a specific subgroup of individuals. The most appropriate methodology will depend on a number of considerations relating to the application process, the fulfilment of the eligibility conditions and the receipt of treatment (the AtW journey, from eligibility to treatment, is illustrated in Figure 3). As detailed below, a number of relevant questions concerning these aspects are still open, and would therefore require further exploration to inform the AtW impact analysis.

Figure 3: Access to Work stages: from eligibility to treatment



5.2.1 The application process

The first important aspect concerns the extent to which eligible individuals (group A) make an AtW application. This is important because eligible individuals who do not apply (group C) are the main candidates for inclusion in the comparison group to be used to estimate impacts. How comparable this group is to AtW recipients who receive treatment will determine the feasibility (and robustness) of estimating the AtW impact. At the time of writing, little is known about eligible individuals who could have made an application but did not do so, and therefore we recommend conducting some research to explore who they are and their likely number. For example, the longitudinal surveys discussed in Section 5.1, below, could be used to identify issues around health related declines among workers and explore to what extent such transitions precede job-loss.

5.2.2 The fulfilment of the eligibility conditions and receipt of treatment

There is reason to believe that the potential pool of eligible AtW non-recipients is large (Section 4.3) and the sample size for this group for the evaluation should be reasonably large. This is because selection into treatment may be driven by specific applicants' characteristics and therefore the pool of potential comparators should be comprehensive enough for the evaluator to be able to choose the best matches for the treated group. That said, it is possible to undertake matching analysis using comparison groups that are smaller than the treatment group. However, there are

Feasibility of evaluating the impact of the Access to Work programme

two drawbacks with this approach. First, there is an increased chance that common support is not achieved. That is, not all people in the treatment group will have a counterpart in the comparison group. This would mean that the impact effect could not be calculated for all AtW recipients. Second, a smaller sample size typically will result in less precise estimates of the impact effect.

In principle, it would be possible to define an Average Treatment effect on the Treated (ATT) using Group D. However, to do so successfully would imply being able to match to an equivalent counterpart subset of Group C (eligible non-recipients). It would be useful to have a better understanding of how Groups D and E vary in order to have greater confidence that any ATT matching Groups D and C are not confounded by Group E's counterpart under C. Similarly, it would be possible to try to find separate matches for Groups F and G among the Group C counterparts. However, in practice, this would be very difficult, and perhaps impossible.

In addition to calculating the ATT through matching the comparison group to the treatment group, we can calculate the Average Treatment effect on the Untreated (ATU), through matching the treatment group to the comparison group. Appropriately averaging the ATT and ATU would give us the Average Treatment Effect (ATE) over the population of AtW eligible workers.

The selection into different treatment subgroups is likely to be the result of a number of choices faced by individuals (and caseworkers), especially given the length and complexity of the application/participation process outlined above. Consequently, it would be useful to have a better idea of the characteristics of the different applicant subgroups to help better understand what is feasible when attempting to match to their counterparts under Group C (eligible non-applicants).

Estimation of impacts for specific subsets of the treated group (e.g. Local Average Treatment Effects (LATE) resulting from the application of the instrumental variable approach) is unlikely to be feasible. A valid instrument to calculate LATE must be associated with the outcome only through its association with selection into the treatment status group. In addition, the instrument needs to be independent of other factors affecting outcomes. These are challenging requirements for AtW and no obvious candidates emerge. Given that the selection process into AtW is based on the assessment of multiple, complex eligibility conditions, a regression discontinuity approach, which requires a cut-off point on a single decision variable, is not deemed feasible. Finally, a Difference-in-Differences (DID) approach is also excluded because it is unlikely that the pre- and post-AtW periods will be comparable (e.g. because the same circumstances leading someone to make an AtW claim may not be observed later, when receiving support). The DID approach would also require that employment outcomes observed under AtW could be compared to employment outcomes before the implementation of AtW. In this case, the pre-treatment would need to precede the implementation of AtW in 1994. Further, we would need to be able to split the pre-treatment AtW eligible population into AtW recipients and non-recipients, which is clearly impossible.

5.2.3 Statistical matching and challenges

We conclude that statistical matching is the only viable option that might enable robust estimation of the AtW impact effect. However, there are many challenges to overcome. The technical issues and assumptions underlying matching will not be rehearsed here but it is worth noting that a key requirement for matching is the conditional independence assumption (CIA). In practical terms, this requires the two groups (AtW recipients and eligible non-recipients) to be equivalent after controlling for the variables included in the matching model. More formally, potential outcomes are assumed to be independent of treatment status. Given the complexity of AtW, this is a demanding assumption to meet. In addition, it is not possible to test that it has been met in full, though checking covariate balance after matching is an important step.

In principle, the impact of the AtW can indicate:

- Changes in the proportion of AtW recipients who are still in work, gain a promotion or experience a health improvement over a fixed period of time (e.g. 3, 6 or 12 months after starting the employment spell). For employment retention, changes in the average number of days worked over the total number of workable days within the same periods can also be explored.
- Hazard rate over variable time periods (allows for censoring issues related to different observed employment spells) allowing survival curves and percentile estimates of duration in state.

In practice, the most likely approach to estimating AtW impacts will entail either a mean comparison of outcomes (after having implemented some form of matching, or within a regression framework), rather than using survival analysis techniques. The use of survival analysis within a propensity score matching framework is a relatively recent development; whereas testing the difference between treatment and control group means is well established. Moreover, given the anticipated follow-up period for the outcome measure is expected to be of a relatively short duration (12 months or less), the advantages of estimating survival curves are less likely to be observed with shorter time frames.

5.3 Measuring the counterfactual

The key concern of any impact evaluation of the AtW will be identifying appropriate people from the pool of all employees who did not participate in the programme, and whose outcomes can be used to provide a proxy for the counterfactual. As explained before, those who are eligible for AtW but do not make an application are the preferred candidates to start with for the purpose of selecting the comparison group. It seems highly likely that such a group exists. For example, a recent Commons

Feasibility of evaluating the impact of the Access to Work programme

brief¹³ reported around 3.5 million people in work with disabilities between April and June 2017. This number obviously far exceeds the recipient sizes reported in Section 2, above. It is not known how many of these 3.5 million would meet the AtW eligibility conditions, but it seems likely that the potential pool of non-eligible recipients is large.

Understanding the triggering events behind making a claim for AtW is important because this will help with the matching process. One trigger might simply reflect a change in awareness of AtW, i.e. workers already meeting eligibility conditions becoming aware of the presence of AtW. Another set of triggers likely relate to changed circumstances among people in work. In particular, changed health circumstances would be expected to play a role, but other circumstances might also be influential, e.g. people moving house and becoming eligible for help with travel to work.

In practice, we would expect employment and health history to be primary in determining who proceeds to the application stage. For example, those with unstable employment patterns prior to claiming may be more inclined to seek AtW support (reflecting financial uncertainty), likewise those with a health condition which has been worsening over time. However, many other factors are likely to play a role in the selection of treated individuals, and will therefore have to be accounted for either in the matching process as covariates or as stratifiers within which matching is undertaken. We are interested in identifying variables which are both determinants of selection into applying and receiving AtW support and prognostic of outcomes (throughout this report we refer to these variables as 'control' variables).

In order to choose appropriate control variables for matching there needs to be a better understanding of the trajectories which can potentially lead to triggering a successful AtW claim and the subsequent outcome trajectory. Various individual differences may be influential here and can mediate or moderate outcomes and the decision to claim in the treatment context. For example, some employers (e.g. large and public sector businesses) may be more willing or able to provide support irrespective of AtW. Also influential may be employee's age, gender, ethnicity and family composition. The health condition(s), and their recurrence and severity may also be important determinants of the support received. Although the final list of control variables will have to be finalised, we would expect the following aspects to be important:

- Work history
- Health history
- Health conditions (main impairment and other concurrent conditions), including whether recurrent and their severity
- Demographics (e.g., age, gender, ethnicity and family composition)
- Employer (size and sector)
- Whether paid- or self-employed
- Nature of current job (sedentary, on-location and ease of access)

¹³ [House of Commons Library brief](#)

- Travel to work distance and mode of transport
- Family composition and financial support
- Education

Identifying the appropriate variables is one challenge; another is having measures of control variables available at the appropriate point in time. It is vital that matching takes place using control data obtained prior to the intervention. Given that AtW is ongoing, finding an appropriate match when a sudden health shock moves someone in work into eligibility requires having data on the timing of the health shock; hence the reason for including health history in the matching model. This requirement for health history, and correct timings of health changes, not only puts large demands on the data (and respondent) requirements, it also raises the question of how we can know if the health shock was relatively minor or sufficiently large to threaten the person's ability to continue in work. It is important to understand how we can construct collection instruments to reliably collect such data.

5.4 Observation window for the study

An AtW impact study requires longitudinal data observed across multiple points in time to capture appropriate pre-treatment variables for matching, identification variables for AtW eligibility and/or receipt of support, and outcome measures. We are also concerned with establishing the minimum number of observation points (or time periods) per person, the unit of measurement of duration (e.g. day, month or year) and whether a sliding window of observation over calendar time should be used to measure outcomes (e.g. a window start defined by the individual's AtW approval/provision date).

At a minimum, one would need to identify three observation points, t_1 , t_2 and t_3 (with $t_1 < t_2 < t_3$), such that the control variables are observed at t_1 (before treatment), treatment status (date of AtW approval¹⁴) is observed at t_2 and outcomes at t_3 (post-treatment). In addition, we would be looking to collect event history data associated with each observation point in time, either from administrative and/or survey recall data.

A sliding window of observation may imply that each point in time differs across individuals. Outcomes may be available at multiple time points after treatment, in which case the evaluator would be able to observe an impact trajectory rather than a single outcome point estimate.

¹⁴ As discussed above, it would be useful to distinguish between new, repeat and renewal claims.

6 Data options

The quality of an Access to Work (AtW) impact evaluation will largely depend upon data availability. The crucial criteria that potential datasets must fulfil to enable a quality study are:

- Sufficient numbers of AtW recipients and non-recipients to detect a substantively meaningful effect size
- The breadth of subject domain coverage required to capture AtW triggering events and to cover the heterogeneity of recipients and their potential comparators
- Indicators of treatment status and measures of relevant outcomes
- A sufficiently large window of observation to enable identification of appropriate variables at the pre-intervention matching stage, the intervention time point and the post-intervention outcome stage (see Section 4.3)

This is a challenging set of criteria for any one dataset to provide. Moreover, the relatively small number of AtW recipients in the general population means that general purpose household surveys will have to be of a prohibitive, substantial size to identify sufficient numbers of AtW recipients to detect meaningful impacts. With around 25,000 AtW recipients in 2016/17 and around 31.8 million people in UK employment, AtW recipients represent 0.08% of employees. In fact, from a general purpose household survey, we would want a sample of over 0.5 million households to find around 500 AtW recipients using simple random sampling.

The two main sources of secondary data considered by this feasibility study are administrative data and survey data.

6.1 Secondary survey data

Consideration of sample size requirements alone effectively rules out the vast majority of secondary social survey data for the purposes of evaluating the AtW impact. However, three survey data sources are worth considering: Understanding Society (USoc), the Labour Force Survey (LFS) and the Annual Population Survey (APS). However, it is unlikely that any of these include a reliable measure of AtW receipt, and therefore they would necessitate linkage to Department for Work and Pensions (DWP) administrative data on AtW recipients. Issues related to obtaining linkage consent from survey respondents would need to be worked through before such an exercise could be undertaken, and survey managers might need reassurance that the addition of such a consent request to the LFS/APS would not be

Feasibility of evaluating the impact of the Access to Work programme

detrimental to response rates. Currently, there is a linkage agreement with DWP and USoc respondents but not with LFS/APS. ONS is exploring possible linkages between survey and administrative data under its Transformations programme, but this is still ongoing and cannot inform this report.

USoc is a general purpose longitudinal survey with around 27,000 households. We might therefore expect to find around 25 AtW recipients in any one year, many of whom may be in the stock rather than the inflow of new recipients. Clearly, even combining longitudinal observation windows across different years will not deliver a large enough sample for a meaningful evaluation.

The LFS is a rotating panel survey with each panel of respondents interviewed up to five times over five quarters. From a longitudinal perspective, the survey exceeds the minimum three data time points requirement described above. It has a sample of around 39,000 households a quarter¹⁵ and therefore it is expected to produce around 36 AtW recipients in any one quarter; again mixing stock and flow AtW recipients. Intuitively, it might seem appropriate to pool quarters of data to increase the number of AtW recipients but the rotating nature of the panel means that only around just over one-fifth of the survey each quarter provides a new sample. Consequently, a pooling strategy would need to take the panel design into account. For example, assuming one fifth (seven) of the 36 AtW recipients in the stock at the beginning of the observation period will be found in each new quarter, pooling seven quarters would give 49 AtW recipients to add to the 36, over a two year period. It would require eight years of data just to get a sample size of around 250 AtW recipients. Clearly, sample size will be an issue even with pooling over quarters.

The APS combines the LFS with local area sample boosts to get a sample of around 150,000 households per annum. We might therefore expect to obtain around 300 AtW recipients in a year. However, whilst the APS is another rotating panel, it comprises the LFS regular sample on a quarterly rotating basis and the LFS boost samples, which rotate on an annual basis. The LFS and APS both provide the appropriate longitudinal data to cover the minimum three observation periods but would need to be combined in a complicated way and would still require many years of data pooling to get a usable sample size. It is possible though that, should data linkage to DWP AtW records prove possible, the APS might provide an opportunity to explore the potential size and characteristics of the AtW eligible, non-claiming population.

Even though the potential of USoc and the LFS is limited for primary analysis in an impact evaluation study of AtW, it is worth noting that they have some potential for helping inform the evaluation. In addition to their use for informing health and work transitions, there is some scope for follow-up research with targeted sub-groups of survey participants. However, this scope is limited. USoc is concerned with the potential effects on response for the next survey wave of interviews, so requires a full consideration and approval process. In addition, follow-up studies are usually limited

¹⁵ [ONS Labour Force Survey User Guide](#)

to qualitative studies. ONS run an omnibus survey as a follow-up to the LFS, but the sample size is relatively small.

The DWP Life Opportunities Survey (LOS) was a three-wave survey based on an achieved sample of over 19,000 households at Wave One, with two further waves of follow up for people with disabilities at Waves Two and Three. Although LOS may not be usable directly in the AtW evaluation, it may provide a valuable source of information for understanding health and work transitions, which might help to inform understanding of AtW more generally.

Given the requirement for a large sample size of AtW recipients and a wide range of longitudinal data covering people from a variety of backgrounds, a primary data collection exercise has some theoretical advantages. However, on the downside, it is likely to be very expensive and therefore careful consideration should be given to a number of aspects. These include, for example, the choice of variables to be collected, the number and type of individuals to be targeted, as well as sampling and data collection strategies for AtW recipients and non-recipients. In principle, collecting data from AtW recipients is comparatively straightforward because they can readily be sampled from AtW administrative data. The real challenge lies with identifying the eligible AtW non-recipient population.

6.2 Administrative data

AtW recipients can be identified directly from DWP administrative records. Moreover, it will be possible to use the HMRC employment records held by DWP as part of the Work and Pensions Longitudinal Study (WPLS) to provide retention outcome data, although there may be some limitations to these data (e.g. self-employment). It will also be possible to use both tax and benefit records to provide work and benefit histories prior to AtW receipt to help select the comparison group. However, in general, it is anticipated that administrative sources will suffer from a major lack of data to define long-term health conditions (though this may be available for the minority group coming from disability benefits). This lack of health data will be particularly disadvantageous for identifying the comparison group.

In general, there will be limited data available to undertake matching using the variables described above as controls. Moreover, the availability of administrative data will be patchy, depending upon a person's work and benefit history. Consequently, it is not envisaged that administrative data can be used to investigate all AtW recipients, though it may be possible to use the data for some subgroups. It may prove possible to combine data from mixed survey and administrative sources. Certainly, using AtW recipient data would be desirable to identify AtW recipients, even within a survey context. However, comparison groups may have to be identified using mixed administrative and survey data giving rise to the possibility that measurement instruments are confounded across treatment and control groups. If

such an approach is taken, it will raise the potential for bias caused through differential measurement error.

Overall, an AtW evaluation focusing on all employees receiving AtW would seem likely to require a bespoke survey, perhaps sampling recipients from claimant data. However, it is currently not clear how best to sample the pool of potential eligible AtW people to form a comparison group. It seems likely that such a group of people exist (Section 4.3) but they would need to be identified, probably through a screening of a larger sample of people in work (see below).

6.3 Primary data collection

In considering a primary data collection exercise, the aim of the study will be to estimate an impact using a matched groups design with a counterfactual representing the absence of AtW completely. This study firstly considers sample size issues, followed by potential sample designs and observation window periods. This report has already discussed some of the subject domain variables required in the collection instrument, so will not be repeated in detail here, although issues with different modes of collection will briefly be considered.

6.3.1 Calculating sample size

A key question for any evaluation study is what sample size is required. Answering this question for a randomised control trial (RCT) involves a number of assumptions and decisions, but is more complex for a matched design because of some of the uncertainty around potential matching options. This report considers first the RCT requirements and then expands upon these with implications arising from possible matching options.

Any conventional sample size calculation for an RCT requires decisions about the significance level of the statistical test for the impact, i.e. the Type I error rate (the acceptable risk of a 'false positive') and the direction of the test. The conventional level is a Type I error rate of five per cent and a two-tailed test is usually adopted. Similarly, convention often sets the statistical power of a test (relating to the risk of a 'false negative' result) to be 80 per cent. These are arbitrary decisions and reflect the risk appetite of the evaluation designers and can be changed to levels which the evaluators deem to be appropriate for any given study. However, the values presented here are those often used and provide an appropriate starting point for the purpose of illustration and preliminary planning. Another key factor is the minimum detectable effect (MDE) of the impact. The MDE is set in advance by the evaluator and, ideally, would reflect a threshold beyond which the returns from the policy exceed the costs of set-up and maintenance. Finally, the natural variability (variance) of the outcome itself is required in the calculation.

Under the assumption of simple random sampling, the basic formula for estimating sample size (for a single group) is:

$$n = \sqrt{\frac{(z_{(1-\alpha/2)} + z_{(1-\beta)})^2 (2\sigma^2)}{d^2}}$$

Where z refers to the normal score of the value for the Type I error (α) and Type II error (β), d refers to the MDE and σ^2 refers to the variance of the outcome, which can be calculated as $\theta(1-\theta)$, where θ is the baseline proportion of 'success', e.g. the proportion of people employed within six months of an AtW award. In practice, the sample size may need to be adjusted for the survey design, where stratification will tend to reduce the required sample size, but clustering will increase it. In addition, any weighting applied to the analysis, e.g. inverse propensity score weighting, will tend to increase the required sample size. However, controlling for correlates of the outcome within an analysis of covariance framework can reduce the sample size requirement. McConnell and Vera-Hernandez (2015) provide an accessible introduction to sample size calculation.¹⁶

It is important to determine the necessary input parameters in advance, but assuming conventional values for Types I and II errors and with a 50% baseline and an MDE of five percentage points, would give a required achieved sample size of around 1600 per group. This initial estimate would have to be adjusted for anticipated non-response levels along with the design, weighting and modelling factors discussed above.

6.3.2 Sample design

A dual frame approach seems sensible given that we are interested in identifying AtW recipients and DWP records can be used to identify them directly. Conversely, identifying a comparison group will be much more difficult given that there is no sampling frame that permits identification directly. Consequently, a two-phase survey design seems most appropriate for the comparison group, with Phase I acting as a screener to identify people meeting the characteristics associated with AtW recipients. Phase II will then collect the further data needed for detailed matching and analysis.

It is envisaged that the Phase I sample will be drawn from a frame created from HMRC PAYE records in order to restrict the first phase sample to people in employment. The sampling fraction will have to account for the expected prevalence of potentially eligible non-AtW recipients in the working population, and then be uprated to account for non-response, and adjusted for the desired ratio of control to

¹⁶ See also Kalton et al (2005), who, unlike McConnell and Vera-Hernandez, also cover the contribution of the survey weights to the design effect impacting upon the sample size calculation. Though Kalton et al discuss weighting in the context of population estimation and non-response, the same principles apply to inverse propensity score weighting, where this technique is used to estimate the counterfactual.

treatment sample size and any sample design/weighting factors. Currently, the size of the potentially eligible AtW recipient population is not known, although it seems that it may potentially be large (Section 4.3). However, further analysis of existing data sources would be useful to get a better prevalence estimate prior to estimating sampling fractions for a primary survey data collection.

Ideally, the sample will be designed to reduce the variance of the estimator, but the mode of the data collection will influence the cost and the associated data collection methodology. Typically, stratification is used to reduce the variance of the impact estimator and it may be possible to use stratification appropriately here; though it will be necessary to undertake some prior work on determining how best to use any survey stratifiers in the matching procedure (see e.g. King and Nielsen, 2016). A face-to-face data collection mode is expensive and for that reason, observations are often clustered geographically. This process may have implications for sample size estimation. For example, if the survey weights, based on the inverse of selection probabilities, are to be used in the impact estimation, e.g. a Population Average Treatment Effect, then survey weights may increase the required sample size through increasing the standard error of the estimator. Consequently, it is proposed that clustering only be used within the context of a face-to-face survey design. Stratification is mostly beneficial and further research is recommended to identify those variables available on the sampling frames that are strongly related to outcomes to use to stratify the sample.

6.3.3 Data collection mode

In order to achieve a suitable sample size of eligible AtW non-recipients there are a number of steps that are required to filter out those who are ineligible. Given that we have little a-priori data to help with the filtering, this is both a challenging and potentially expensive task.

Our best current upper limit estimated size of the non-eligible AtW population is around 3.5 million from around 30 million employed. Consequently, we need to select around 10 working people to find a single worker with a disability. Moreover, because we do not know the proportion of working people with health problems who meet the AtW eligibility criteria, we do not know how many of those workers with a disability need to be filtered out to find an AtW counterpart for the comparison group. If, for example, half of the 3.5 million workers with a disability met the AtW criteria then we would need to select around 17 people for each AtW eligible. However, if only one in ten workers with disabilities met the AtW eligibility conditions then we would expect to select around 86 people to find an AtW eligible person. Add in the survey non-response rate and this figure increases further. For example, assuming 50 per cent for both the eligibility and non-response rates would require us to sample 72,000 workers to find 2,000 AtW eligible workers. If we decrease those rates to 20 per cent, the initial required screening sample size would be around 450,000.

The above discussion demonstrates that a large Phase I screening sample will be required to detect the appropriate number of comparison group respondents. It is important first to estimate the desired sampling fraction, as discussed above, before

Feasibility of evaluating the impact of the Access to Work programme

calculating the cost of different modes of data collection and the likely response rates. Face-to-face is the most expensive data collection mode but is likely to have the highest response rates. However, it may be more attractive to use a web data collection mode, which may only have a one in five response, but will be much cheaper than sending interviewers out to many addresses.

The key attraction of self-completion, in the form of a web survey, is the removal of interviewer costs, which are incurred by both face-to-face and telephone surveys. The key costs are the fixed cost of setting up the data collection instrument and data processing; and the variable cost of letters of invitation to participate. From a statistical perspective, the key risk is an unrepresentative sample arising from low response. However, a low response does not guarantee an unrepresentative sample. Moreover, if the key estimator is the treatment on the treated then from the perspective of the comparison group survey, the key aim is to ensure common support across the treatment propensity score rather than a representative sample. If an impact estimate is required from the sample to the wider population of AtW recipients or the eligible population, variables drawn from the HMRC tax data could be used in calibration weighting of the sample data. Calibration can readjust sample imbalances to reflect distributions in the wider population, i.e. it can adjust for some sample to population imbalances which give rise to unrepresentativeness.

For the AtW population, it may be more desirable to ensure a representative sample than is required for the comparison group. In this case, a face-to-face survey may be practicable, given the relatively low sample size requirements enabled by direct targeting of AtW recipients compared to the untargeted Phase 1 sampling requirements for the comparison group. Nevertheless, it is not anticipated that there would be much opportunity for the geographic clustering of AtW recipient addresses, which would lead to comparatively high interviewer travel costs. Consequently, it would be desirable to consider the cost-benefit trade-offs between web and face-to-face modes and to balance these with the potential risks to the degree of representativeness of the likely achieved samples. In such a mixed-mode design, it is possible that systematic changes to interview mode could systematically affect survey responses in a way that confounds estimation of the impact effect. If a mixed-mode study were to be considered, we recommend first piloting the questions under different interview modes prior to the evaluation. This would permit a more informed judgement to be made concerning the risks to quality against the costs of alternative options.

7 Exploiting changes in Access to Work conditions

The focus of this report has been given over to the potential for estimating an impact against a counterfactual of a business as usual scenario with no Access to Work (AtW) availability because this gives the most authoritative estimate of an AtW impact. There have been a number of changes in AtW conditions over the years and the following section considers what these changes can tell us in an impact evaluation context.

7.1 The introduction of Personal Independence Payments

Recently, Personal Independence Payments (PIP) were introduced to replace Disability Living Allowance (DLA). This is a phased replacement with new claims having to be made for PIP, but a staggered approach to moving the stock of DLA claims gradually across to PIP, alongside an assessment to meet the PIP eligibility criteria.

Department for Work and Pensions (DWP) already hold data on DLA, PIP and AtW and could link to HMRC tax records in order to extract the appropriate data both for matching and for outcomes. There are indications the population subgroup sizes would be sufficiently large in covering AtW recipients claiming PIP/DLA to undertake a relatively robust estimation. However, DWP need to do further work to finalise and quality assure these data before recommendations can be made.

One potential route of enquiry is to match PIP/DLA recipients in work receiving AtW to their counterparts who are in work but not claiming AtW. Work-related outcomes could then be obtained through linkage to HMRC tax data. Matching would take advantage particularly of disability/health data from DWP records; but whilst potentially these data would have good coverage from a historical perspective, they might only cover a subset of health issues. However, there would be a wealth of data available from the PIP/DLA application process available for matching.

The move to PIP has led to some former DLA claimants no longer receiving the support they had under DLA. It may also be advantageous to explore the extent to which AtW take-up has risen as a result of withdrawing DLA. It is important to note that former DLA recipients losing PIP are likely to differ from those who get PIP in ways that may affect their ability to work. Consequently, a simple group comparison is unlikely to be sufficient and matching may be required to make a robust estimate.

7.2 Selective marketing of Access to Work

One promising area is the possibility of marketing and promoting AtW to encourage take-up of work among the potentially eligible AtW population who are out of work. Ideally this design would involve a randomised control trial where the treatment group received increased awareness treatment and the control group received usual advice and guidance.

There are many different approaches to increasing awareness of any service and it is not immediately clear which of these would be most appropriate. However, it would be possible to focus on people with long-term health issues undergoing the Adult Improving Access to Psychological Therapies (IAPT) Programme. Although this is a select subgroup of the potentially eligible AtW population, it offers opportunities to assess the extent to which awareness and promotion of AtW can improve take-up among a relatively well-defined group. In principle, it would also be possible to explore retention and advancement, but the practicality of this would depend upon the numbers of people both undergoing IAPT and being encouraged into work¹⁷ during the process. The impact effect for retention/advancement would then be a function first of increased take-up followed by retention/advancement. If take-up is low, then this would lower the potential for detecting a retention/advancement effect.

It would be possible to generalise the awareness campaign approach to wider groups using different approaches to promotion (household leaflets, radio broadcasts etc.) However, risks of contamination would increase with probable unknown numbers of the control group inadvertently coming into contact with the treatment.

¹⁷ When considering retention and advancement given the random assignment, we could also include any current stock of AtW recipients undergoing IAPT in the analysis.

8 Gaps and further research

This report has outlined the key challenges that the Access to Work (AtW) programme is likely to face in designing an impact evaluation study and choosing among different estimation methods. There are two key challenges for the evaluation. The first challenge is the lack of appropriate existing data sources, especially to create an appropriate counterfactual group. The second arises from an incomplete understanding of a number of relevant issues concerning the size of the comparison group and the selection processes both on behalf of the choices to be made by the potential recipient and decisions made by AtW advisors and employers. Given the likely high cost of a bespoke survey, further research is recommended to inform these issues before taking a decision on the desirability of conducting an impact evaluation.

Department for Work and Pensions (DWP) has recently commissioned qualitative research with AtW recipients, employers and assessors, findings from which have been published alongside this report. This qualitative research delivers insights into, amongst other things, what triggers applications. Findings will help inform a number of points recommended below.

The main recommendations for future research stemming from this study are as follows:

- We still do not know enough about the different stages of the AtW applicant's journey to be able to understand the composition and circumstances surrounding the subjects of interest (eligible individuals, successful applicants and AtW recipients). This implies that the treatment parameter to be estimated cannot be defined (e.g. the Average Treatment Effect (ATE) or the Average Treatment effect on the Treated (ATT)) and, consequently, it is not currently possible to identify the most appropriate methodology to estimate the impact of AtW.
- It seems plausible to assume that selection into the different stages of the customer journey is likely to result in treated and eligible individuals being two different populations, in which case some form of matching estimation approach would be envisaged to estimate an ATT. However, much has yet to be learned about AtW recipients and their application/selection process before one can commit to any specific impact estimation approach. Future research should focus on understanding the circumstances and dynamics that triggered an AtW claim, as these will suggest what the potential comparator group (if it exists) and its size will have to look like for an impact assessment of the AtW to be robust.
- Any AtW impact evaluation will have to identify the programme's additionality, that is, what AtW provides to the individuals supported that they would not get

otherwise (the 'business as usual' scenario). In doing so, controlling for the existence of reasonable adjustments by employers for the individuals under study (especially, potential comparators) will prove particularly challenging, with a possible risk to mis-state the real impact of AtW. It is not clear what constitutes such adjustments and, to the extent of our knowledge, no existing data capture this aspect directly. A recommendation would be to explore caseworkers' subjectivity in assessing AtW claims. This might provide some insight into variables that could proxy reasonable adjustments and, more generally, employers' support for both AtW and non-AtW recipients.

- No single existing data source (or combination of sources) is likely to provide the breadth and depth needed to cover the heterogeneity of recipients and their potential comparators, and provide sufficient sample sizes to detect substantively meaningful AtW impacts. Administrative data can certainly be used to identify eligible/treated individuals and provide their work and benefit history but the detail required for impact evaluation is likely to be available only for specific subgroups of recipients. Existing surveys are unlikely to provide sufficient sample sizes but could nevertheless be helpful to explore the potential size and characteristics of the AtW eligible, non-claiming population, and may provide a valuable source of information for better understanding health and work transitions.
- Considering the requirement for a large sample size of AtW recipients and a wide range of longitudinal data covering people from a variety of backgrounds, a primary data collection exercise is likely to be the way forward. However, given the high costs expected, several aspects should carefully be considered prior to committing to its implementation. These include, for example, the choice of variables to be collected, the number and type of individuals to be targeted, as well as sampling and data collection strategies for AtW recipients and non-recipients.
- The results from DWP's current qualitative research on AtW should be compared to the issues discussed in this paper to establish the extent to which knowledge gaps have been filled and those that remain and require further work.
- Further quantitative research should be undertaken on AtW applicants to better understand the size of the unsuccessful AtW applicant groups, reason for non-awards, and outcomes of assessment awards.
- More work is required on how best to identify AtW eligible non-recipients from the more general population of workers with disabilities and long-term health conditions. Primarily, this will require secondary analysis of existing datasets but may be supplemented by research with caseworkers using scenarios based upon data which can be collected via surveys.

9 Appendix: Bibliography of Previous Access to Work Research

Aston, J. (2009). Evaluation of Access to Work: Individual Budget Pilot Strand. DWP RR 620.

Beinart. (1997). The Access to Work programme: further analysis of data from the 1995 surveys of ATW recipients and their employers; SCPR report for the Employment Service

Beinart, et al (1996). The Access to Work programme: a survey of recipients, employers, Employment Service managers and staff, Report for Employment Service

Dewson, S., Fearn, H. and Williams, C. (2009). Evaluation of Access to Work: Ministerial Government Strand; DWP RR 621

Dewson, S., Hill, D., Meager, N. and Willison R. (2009). Evaluation of AtW: Core Evaluation, DWP RR 619

Hillage, J. et al (1998). Evaluation of Access to Work, Report for Employment Service

Hirst, M. (2001). Evaluating Access to Work and Workstep, – unpublished feasibility study received by DWP, Social Policy Research Unit, Uni of York

Melville, et al (2015). Access to Work: Cost Benefit Analysis, Centre for Economic and Social Inclusion – commissioned by RNIB

Sayce (2011) Getting in, staying in and getting on DWP

Thornton, and Corden, A. (2002). Evaluating the Impact of AtW – a Case Study Approach, Social Policy Research Unit, Uni of York

Thornton, et al (2001). Users' views of Access to Work: final report of a study for the Employment Service; 1, ES RR 72

10 References

Holland, P. W. (1986). Statistics and Casual Inference. *Journal of the American Statistical Association*, Volume 81(396), p.945-960.

Kalton, G., Brick, J. M. and Le, T. (2005). Estimating components of design effects for use in sample design. In: *UNSD, Household Sample Surveys in Developing and Transition Countries*, p.95-121. Available at:

King, G. and Nielsen, R. (2016). Why Propensity Score Should Not Be Used For Matching. Working Paper

McConnell, B. and Vera-Hernandez, M. (2015). Going beyond simple sample size calculations: a practitioner's guide, IFS Working Paper 15/17, p1-53.