# NatCen

**Social Research** that works for society

IER

# Labour Market Enforcement Evaluation

## Scoping Study

**Authors:** Robert Wishart, Wil Hunt, Nilufer Rahim, Peter Elias, Chris Grollman
**Date:** July 2019
**Prepared for:** The Office for the Director of Labour Market Enforcement

# Contents

## Tables

## Figures

# Summary

This scoping study was commissioned to identify the optimal evaluation framework for assessing the impact of the HMRC National Minimum Wage (NMW) team; the Gangmasters and Labour Abuse Authority (GLAA) and the Employment Agency Standards (EAS) Inspectorate, who work to tackle labour market exploitation.

This scoping exercise has found that the broad coverage and complexity of labour market enforcement and the overlapping powers and responsibilities of each of these bodies makes **a single evaluation encompassing all aspects of enforcement unfeasible at this time**. In addition, the logic model presented in this report identifies different causal chains leading to five distinct outcomes. The optimal research design and available data to assess the achievement of these outcomes will vary between each of the three enforcement bodies.

The study has also found that **some outcomes are not suitable for impact evaluation methods attempting to make a causal inference**. An evaluation framework based on a **'contribution analysis' design offers the most appropriate way of understanding the impact of the three bodies**. In this case, contribution analysis may involve the use of several different methods, including experimental and quasi-experimental approaches; secondary analysis of existing data sets; collection of new survey data; and qualitative research, to draw conclusions about effectiveness.

**UK and international research studies draw on a number of techniques to assess labour market enforcement and to measure non-compliance**. The three most common approaches to assessing labour market enforcement that have been robustly implemented to estimate a causal impact are instrumental variables, difference-in-differences and regression discontinuity design. Several studies use other quantitative techniques, which cannot estimate causal impact but can still provide informative evidence about labour market enforcement. Statutory surveys and targeted surveys have been employed to estimate levels of non-compliance with labour market regulations, though some non-compliance is likely to go unmeasured for the former and a suitable sample frame is needed for the latter. Where non-compliance is most likely to affect hard-to-reach groups, targeted research using a more novel methodology such as Respondent Driven Sampling (RDS) may be needed to establish a baseline measure of non-compliance that can be used to monitor progress on an ongoing basis.

Using this evaluation framework and findings from the literature review, **this study identifies a set of key research questions for which robust methodologies could be implemented** to provide credible options for evaluating the impact of the three bodies. Where no robust approach for establishing causal impact can be considered viable, alternative approaches to assess the work of the enforcement bodies are suggested. The research questions are structured by the three work streams detailed in the logic model (public communications; direct action and prevention) and around estimating non-compliance.

**To assess the impact of public communications, it may be possible to implement a randomised controlled trial** to estimate the impact of "nudges", (emails, letters or texts), signposting employers to existing resources to boost voluntary compliance. **Alternatively, a potentially more efficient approach would be quantitative surveys of employers and employees, combined with qualitative research**. This mixed-method approach may be the best way of inferring how the work of the LME bodies affects understanding of labour market regulations.

The main option for adding value to existing data on complaint-led activity (direct action), would likely rely on new data collection. **The best option to implement an impact evaluation of targeted enforcement would involve secondary analysis of existing data sources using difference-in-differences and instrumental variables**. This would likely draw on data from the Annual Survey of Hours and Earnings and the Labour Force Survey.

Assessing the quantitative impact of the LME bodies' prevention work will likely be challenging. **The most beneficial option for further research in this area may be through qualitative case study work, exploring perceptions of how stakeholder engagement and co-design work (such as industry protocols) affect employers' understanding of their responsibilities**.

In order to undertake a full impact evaluation, it would be necessary to baseline the level of non-compliance before and after the intervention. As such, measurements of non-compliance have been considered here. A variety of forms of labour market non-compliance exist, driven by ignorance of relevant legislation on the part of employers and/or workers; collusion between employers and workers; and exploitation of workers by employers. This has implications for how best to measure non-compliance and who best to approach for the required information; employers, employees or both.

Looking at each body in turn, the first of three recommendations is an **online survey of gig economy couriers** about their working hours, pay, national insurance status, costs of maintaining bicycles etc., to indicate the extent to which employers are avoiding paying workers at or above the national minimum wage by engaging them on a self-employed basis. The second is **a survey of agency workers** about their assignments, pay and employers to measure non-compliance with the right to equal pay that an agency worker should receive under the Agency Worker Regulations, **complemented by a survey of employers about the pay of employees doing the same or similar work**. The third and final recommended approach is **a survey of employees, employed by licensed and unlicensed gangmasters, using RDS to identify the extent of compliance with the GLAA's licensing standards**.

The table below gives estimated timelines for the evaluation options described in this report and provides an indication of how expensive each option is expected to be:

| Table 0:1      Estimated costs for research projects | | |
| --- | --- | --- |
| **Approach** | **Estimated timeline from commissioning** | **Estimated Cost** |
| Analysis of NMW and NLW non-compliance in industries targeted by the HMRC NMW team, using DID and IV | 18 months to 2 years | ☐ £ <br> ☒ ££ <br> ☐ £££ <br> ☐ ££££ |
| Difference-in-differences analysis exploring the impact of different powers between UK countries | Not currently possible. Would require prospective data collection. | n/a |
| Randomisation to assess the impact of promotion activities on voluntary compliance<br>   **Administrative data only** | 3 to 4 years | ☐ £ <br> ☒ ££ <br> ☐ £££ <br> ☐ ££££ |

| Table 0:1 | Estimated costs for research projects | |
|---|---|---|
| **With data collection** | 4 to 5 years | ☐ £<br>☐ ££<br>☒ £££<br>☐ ££££ |
| Telephone survey of employers | 9 to 12 months | ☐ £<br>☒ ££<br>☐ £££<br>☐ ££££ |
| Random-probability omnibus survey of individuals | 6 to 9 months | ☒ £<br>☐ ££<br>☐ £££<br>☐ ££££ |
| Depth interviewing: Employers | 6 to 9 months | ☒ £<br>☐ ££<br>☐ £££<br>☐ ££££ |
| Depth interviewing: Individuals | 6 to 9 months | ☒ £<br>☐ ££<br>☐ £££<br>☐ ££££ |
| Case study approach (12 case studies) | 12 to18 months | ☒ £<br>☐ ££<br>☐ £££<br>☐ ££££ |
| Example 1: Gig economy couriers – non-compliance with pay, hours and contractual conditions of employment | 12 months | ☐ £<br>☒ ££<br>☐ £££<br>☐ ££££ |
| Example 2: Agency workers and non-compliance with Agency Worker Regulations | 12 to 18 months | ☐ £<br>☐ ££<br>☐ £££<br>☒ ££££ |
| Example 3: Exploitation of workers covered by Gangmasters | 18 months | ☐ £<br>☐ ££<br>☐ £££<br>☒ ££££ |

# 1  Introduction

This report presents findings from a scoping exercise conducted by NatCen Social Research (NatCen) and the Institute for Employment Research (IER) which aimed to design an evaluation approach to assess the impact of three labour market enforcement bodies: the HMRC National Minimum Wage (NMW) team; the Gangmasters and Labour Abuse Authority (GLAA) and the Employment Agency Standards (EAS) Inspectorate.

## 1.1  Context for the scoping exercise

A key recommendation of the first annual strategy published by the Director of Labour Market Enforcement was to evaluate the impact of three labour market enforcement bodies that fall within the Director's remit. The three bodies are: the HMRC National Minimum Wage (NMW) team; the Gangmasters and Labour Abuse Authority (GLAA) and the Employment Agency Standards (EAS) Inspectorate[1]. The bodies are responsible for enforcing key legislation that is vital to protecting workers' rights.

The NMW team sits within HMRC and enforces the National Minimum and National Living Wage in line with legislation and policy set out by the Department for Business, Energy and Industrial Strategy (BEIS). The GLAA sits within the Home Office and is an arms-length, non-departmental public body that operates a licensing scheme through which employers who produce fresh products (and associated businesses) are assessed to check they meet standards covering health and safety, accommodation, pay transport and training. The EAS is located within BEIS and exists to enforce the provisions of the Employment Agencies Act 1973[2]. Its remit is to ensure that employment agencies and businesses comply with relevant legislation in order to protect the rights of agency workers. Recent government proposals relating to workplace rights may expand the powers and remit of the EAS.

The differing roles of the three agencies and the legislative parameters within which they operate give rise to a variety of forms of non-compliance, including but not restricted to such actions as failure to pay workers at or above the national minimum wage, the operation of unlicensed gangmasters, and failure to establish the correct terms of employment for agency workers. Non-compliance can be classified according to its causes, which are broadly grouped as follows:

- **Ignorance** of relevant legislation on the part of employers and/or workers – possibly linked to the complexity of legislation or the time it takes for information about changes in the legal framework to become widely available

- **Collusion** between employers and workers – in situations where both parties can share in some pecuniary gain achieved via non-compliance

- **Exploitation** of workers by employers – specifically where there is the possibility for employers to exert a degree of coercive control over their employees via threats or sanctions.

While the three bodies have different remits, structures and resources, they all aim to follow the regulators code (based on the Hampton Principles and Macrory review). The regulations reflect the fact that the work of the bodies needs to balance protecting workers' rights with supporting the industries they regulate, operating transparently and carrying out their duties in a proportionate, risk-informed way.

---

[1] The fourth LME body, the Health and Safety Executive, is not considered in this report.
[2] http://www.legislation.gov.uk/ukpga/1973/35

The three regulatory bodies produce statistics, mostly in the form of annual reports, which focus on performance indicators. While this kind of assessment is clearly important in understanding how well the regulators are operating, it does not provide a robust evaluation of their impact or reliable evidence on the scale and nature of labour market exploitation. The purpose of this scoping study therefore is to generate insight into the issues involved in evaluating the regulatory bodies and to provide recommendations on the best approach to evaluation and to measuring non-compliance, taking into account methodological and practical constraints.

## 1.2 Aims and objectives

This scoping study has been commissioned to advise on the evaluation framework that could best be utilised to evaluate the impact of the three labour market enforcement bodies, presenting research and discussion on the possible methodologies, as well as their relative robustness and feasibility in the labour market enforcement context.

As such the primary research questions for this scoping exercise asked:

- How could one evaluate the impact of labour market enforcement, with regard to the NMW team; the GLAA and the EAS Inspectorate?

- What research methods could be utilised to make a more accurate assessment of the scale and nature of non-compliance?

Underpinning these overarching questions, the scoping exercise aimed to determine what research design is recommended to assess the impact of the labour enforcement bodies in the UK and what resources are required to undertake recommended methods. The DLME was also interested in the scalability and replicability of the suggested design and the extent to which it would include flexibility for structural changes in the labour market.

## 1.3 Approach

Taking a theory-based approach, the first stage in deciding which evaluation approach to take was to develop a **logic model** to articulate how LME activities are expected to lead to intended outcomes. Establishing intended outcomes in this way gave us a clearer indication around what research methods could be implemented to explore whether each body is achieving its intended outcomes.

The development of the logic model was informed by eight telephone interviews with key senior stakeholders from each body and sponsoring departments BEIS and the Home Office. The interviews were supplemented with a review of ten relevant documents about the work of the bodies (see Appendix A). An overarching logic model was presented at a workshop consisting of the DLME, LME bodies and their respective sponsoring departments.

The second stage involved a **rapid review of literature** around evaluation frameworks of enforcement bodies, informed by a set of key research questions. Studies included in the review used randomisation or a quasi-experimental design conducive to a causal claim; were conducted in the UK or another country with a suitably developed economy and broadly similar labour market regulations; and were peer-reviewed journal or "grey literature" from a trusted source (see Appendices D and E).

A small number of internal **experts** were consulted on their views of the types of research methods available to measure the impact of enforcement strategies. Informal discussions were held with experts on an ongoing basis. In addition, meetings were

held with each of the LME bodies to seek further detail on the data collected and held by the bodies, and the potential for the data to be used for evaluation purposes.

In order to undertake a full impact evaluation, it would be necessary to baseline the level of non-compliance before and after any intervention. The work undertaken by IER in designing an approach to measuring the scale of non-compliance consisted of a detailed review of academic and policy literature in the UK, but with a key focus on international comparators to learn from international best practice. This work explored existing approaches and methodologies around measuring non-compliance and considered their relative strengths and weaknesses, particularly within the labour market enforcement context, to inform the application of suitable methods to the UK labour market enforcement context.

A final **workshop** was held with the DLME, LME bodies and their sponsoring departments to present findings from the reviews of literature and to discuss a range of potential research questions and corresponding methods that could be employed to build a picture of the impact of the LME bodies, and to understand the scale of non-compliance with labour market legislation in the short, medium and long term.

## 1.4   Potential evaluation methods

For the purpose of this report, impact evaluation is referred to as any quantitative approach that can be used to robustly estimate causal impact on outcomes. Specifically, this refers to methods that establish a **counterfactual**. This means that a comparison group is established to assess the difference in the outcome of an individual or business if they were allocated to an intervention and the outcome if they were not.

Randomisation is perceived by many as the "gold standard" in establishing causal impact. **Randomised controlled trials** (RCTs) have been used extensively in medical research, but in recent decades they have been increasingly used in social sciences, such as education. In an RCT, individuals are randomly allocated to receive an intervention or to receive no intervention[3].

Random allocation is different from random sampling. Random sampling involves randomly selecting individuals within a population and surveying them. Random allocation takes a sample of individuals and randomly assigns a proportion of this sample to receive an intervention, with the remaining individuals assigned to control. Provided that the sample is big enough, the groups should have the same composition.

In some situations, randomisation is not possible, or not desirable. For example, an RCT cannot be implemented in situations where an intervention has already started, or if there are significant ethical implications for conducting an RCT.

In these situations, **quasi-experimental designs** – statistical techniques that construct a counterfactual – can be used to estimate causal impact. There are several statistical methods that are considered quasi-experimental designs. Some of these designs, including real world examples, are included in Table 1:1.

---

[3] In practice, this group often continues to do what they would do in the absence of the intervention, which is referred to as "business as usual".

| Table 1:1 | Quasi-experimental design approaches | |
|---|---|---|
| **Name** | **Definition** | **Example** |
| Regression Discontinuity Design (RDD) | Where an administrative decision is based on a continuous measure, individuals with scores very close to the cut-off can be assumed to be the same, simulating random assignment. There is a trade-off between sample size and bias. Having a wider inclusion criterion increases the number of individuals in the sample, but risks introducing bias if the criteria are too broad, making individuals not comparable. | Pupils sitting the Eleven-plus go to grammar school if they achieve a certain mark. If the pass mark is 80, pupils with a mark of 80-82 could be considered the same as pupils with a mark of 77-79. The impact of grammar schools on educational attainment could then be assessed by comparing these grammar school pupils to comprehensive school pupils. |
| Difference-in-differences (DID) | When two groups exist and the change in an outcome over time has a "common trend", the DID approach can be used. The "common trends" assumption assumes that changes in the achievement of an outcome over time are same for both the "treated" and the "control" group. To estimate the impact, take the change in the difference between the two groups in a time period before the intervention, and after the intervention. | In the U.S. where minimum wages were set at state level, Card & Kreuger (1994) compared the change in employment at fast food restuarants when the minimum wage was increased in New Jersey, but not in Pennsylvania. The employment rates were compared before and after the increase in the minimum wage in both states. |
| Propensity Score Matching (PSM) | If all the relevant characteristics of an individual can be observed, people allocated to receive an intervention are allocated "propensity scores" based on these characteristics. People not in the intervention are also allocated these scores based on their characteristics, and an algorithm matches individuals with similar scores. It is important to establish "common support", where each treated unit has at least one comparison unit with a similar score and ensure that there is no bias on observed covariates. | If a comparison is made between incidence of heart disease and smoking compared smokers and non-smokers, the proportion of smokers with heart disease is much larger than the proportion of non-smokers. However, using propensity score matching removed confounders such as age and socio-economic background, reduces the difference between the two groups, though it is still significant. |

| Table 1:1 | Quasi-experimental design approaches | |
|---|---|---|
| Instrumental Variables (IV) | The instrumental variables approach is implemented when a covariate suffers from "endogeneity" – i.e. it is correlated with the error term and could therefore bias results. If another observed covariate is correlated with the endogenous variable, but not with the error term, it can be used as an *instrumental variable* in the regression model. This approach relies on being able to establish a suitable instrument, which can often be difficult. | Correlation between smoking and poor general health cannot establish a causal claim as smoking may be influenced by other health factors, such as depression, which could affect both smoking and health. An RCT is not ethical in this situation as you cannot force people to smoke, so smoking can be instrumented through the tax rate on tobacco products, which does not affect health directly, but does affect tobacco consumption. Angrist & Kreuger (2001) provide a good summary of studies using IV estimation. |
| Synthetic Control Group | A synthetic control group is used to evaluate the effect of an intervention by constructing a weighted combination of groups used as controls, to which the treatment group is compared. Outcomes can only be assessed at an aggregated (group) level and the synthetic control must have a similar trend in changes in the outcome for a large number of periods before the intervention takes place. | To estimate the impact of reunification on West Germany on economic growth, Abadie et al. (2015) created a synthetic control group. No single country was able to closely approximate the values of economic growth predictors for West Germany before reunification. A weighted average of Austria, the United States, Japan, Switzerland, and the Netherlands provided a very close approximation to West Germany prior to 1990 and could therefore be used as a comparison group against the actual growth of West Germany. |

In practice, sometimes it is not possible to provide causal estimates of impact. Other quantitative approaches can sometimes be used to assess associations between outcomes, but they cannot be used to infer causal impact. Many labour market interventions are too complex to be analysed with an impact evaluation approach given the wide range of possible confounders, or the required data to facilitate that analysis is not available. It is therefore often necessary to triangulate findings of several studies to infer impact using a 'contribution analysis' approach. This can include impact evaluations, assessing trends from administrative data, other quantitative analysis and qualitative evidence from interviews and focus groups. As is highlighted later in this report, it is likely that this type of approach may be the most realistic in assessing the impact of labour market enforcement.

## 1.5   Structure of the report

This report provides a detailed assessment of the feasibility of carrying out an evaluation of the three LME bodies and of measuring the scale of non-compliance. The structure of the report reflects the methodology employed.

- **Chapter 2** presents the logic model designed to articulate the impacts and outcomes of the three bodies

- **Chapter 3** discusses findings from two literature reviews on approaches applied in the UK and more widely to evaluate the impact of labour market enforcement and to explore what methodologies to the measurement of compliance/non-compliance have been adopted elsewhere

- **Chapter 4** explores a set of evaluation research questions and methodologies for answering them

- **Chapter 5** concludes this report by highlighting the recommendations for evaluating the work of the three bodies.

# 2  Theory of Change

A theory-based approach was taken to determine the most appropriate design to conduct a robust evaluation of the three labour market enforcement bodies. A theory-based approach stipulates that all programmes of work have an underlying theory or rationale. Articulating the underlying theory helps to structure how an evaluation can be designed and conducted and helps to establish what the main outcomes are.

Programme theory is articulated by making explicit how a set of activities are expected to lead to the intended results of an intervention. The process of sequencing all components of an intervention forms pathways that link activities to outcomes and ultimate programme goals (or impact) and is typically represented visually in a logic model.

As depicted in Figure 2:1 logic models set out the relationship between the resources available to an organisation (its inputs), the activities it undertakes, the immediate effects of those activities (outputs), the broad goals to which those outputs contribute (outcomes) and the overarching impacts that the organisation seeks to achieve.



Figure 2:1     Kellogg Foundation Logic Model

| Inputs | Activities | Outputs | Outcomes | Impact |
|---|---|---|---|---|
| The resources an intervention has available to direct toward doing the work | The processes, tools, events, technology, and actions that are an intentional part of implementation | The direct products of programme activities | Specific changes in behaviour, knowledge and skills | The fundamental change the intervention seeks to bring about |

*Planned work*                                        *Intended results*

The development of a logic model was the first key stage in this scoping study, to map the roles, remits and responsibilities of the bodies and clearly articulate the links between inputs, activities, outputs and intended outcomes of each body.

This chapter presents a logic model designed to encompass the work of all three LME bodies, the design of which was informed by in-depth interviews with officials from the three bodies and a review of documentation (a full reference list for documents cited in this section can be found in Appendix A). Following this, the chapter discusses contextual factors that are seen to enable or constrain the ability of the LME bodies to achieve the stated outcomes and their overall goals.

## 2.1   Overarching logic model

An overarching logic model for the three LME bodies is shown in Figure 2:2 (a detailed version of the logic model can be found in Appendix B). Each component of the logic model is then described in turn, drawing on information captured in in-depth interviews and a review of documentation. The work of each body shares broadly similar features

and has been summarised into one logic model for the purposes of evaluation design. It is however important to note that the picture is more nuanced in reality and that some aspects of the logic model or strands of work presented within it may apply more to one body than another. 'Worked examples' of how a strand of each body's work flows through the logic model are provided in Appendix C.

Figure 2:2     High Level Logic Model
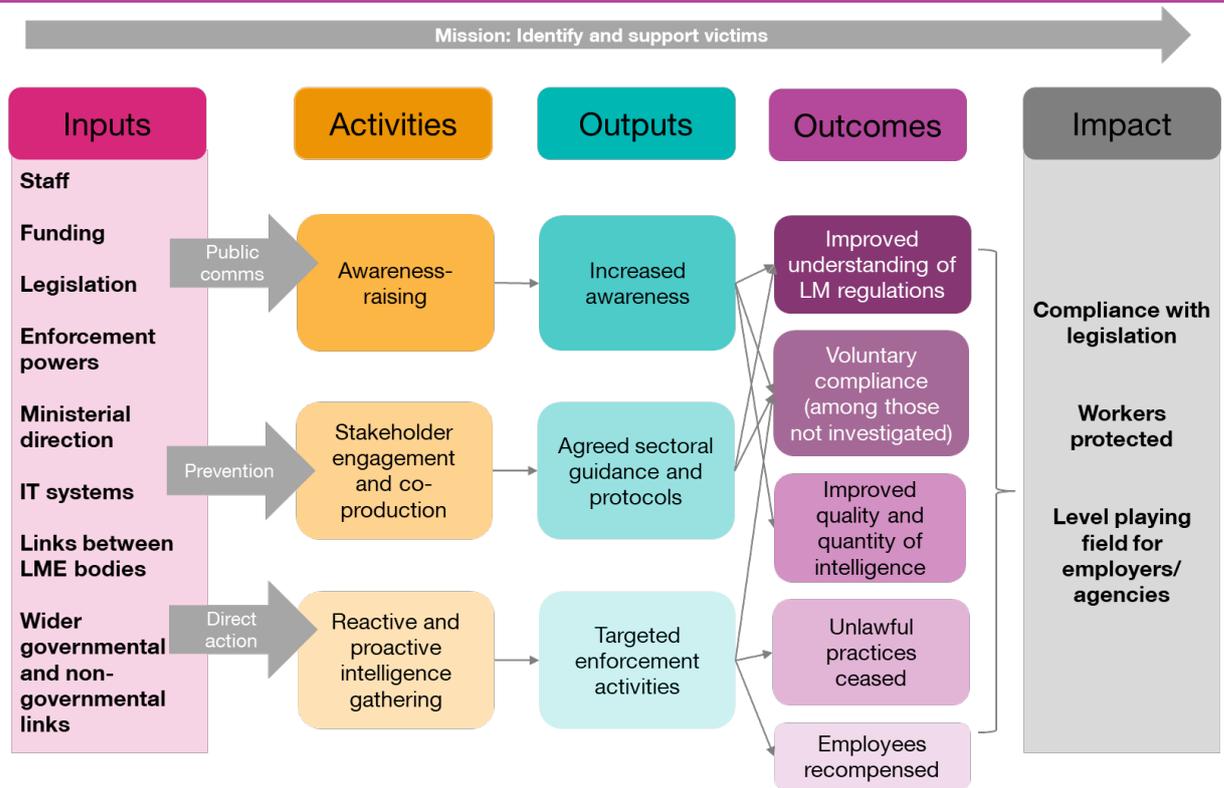


## 2.1.1 Impacts

The overriding mission of the LME agencies is to identify and support victims of labour market infringements. Underpinning this are the aims of achieving compliance with legislation and creating a level playing field for employers/ agencies; where all parties have a fair and equal chance of success.

## 2.1.2 Inputs

The inputs common across all three bodies include staff, funding, legislation, enforcement powers, ministerial direction, IT systems, links between the LME bodies and wider governmental links. Table 2.1 summarises key inputs alongside further information about the three agencies.

**Staff**
The workforce of each body in varies size, with the NMW team being the most substantial. In 2017/18 the NMW team consisted of over 399 members of staff; the GLAA had approximately 104 staff and the EAS, the smallest of the three bodies, had 12 members of staff. At the time of interviews in late 2018 all three agencies had expanded or were in the process of doing so to reflect wider remits prompted through legislative changes.

**Funding**
The NMW team receives £25.3m per year. The GLAA has an annual budget of just over £7m which has increased from £4.8m to reflect a wider remit (see Legislation). EAS receives core funding of £0.5m per year, with an additional payment in 17/18 to enhance its IT infrastructure.

**Legislation:**
Several different statutes form the legislative basis for the three bodies and further statutory changes are set to expand their remits. The legal basis for EAS is set out in the Employment Agencies Act 1973 and the Conduct of Employment Agencies and Employment Businesses Regulations 2003 across Great Britain[4]. The National Minimum Wage Act 1998 established a minimum wage across the UK. Each year, HMRC and BEIS agree a Service Level Agreement under the current legislative framework. The GLAA was established following the 2016 Immigration Act. Before this, it was named the Gangmasters Licensing Authority (GLA), under the Gangmasters (Licensing) Act 2004. The GLAA retains the former GLA function of licensing gangmasters providing labourers for agriculture, gathering shellfish or processing food.

The 2016 Immigration Act introduced the role of Labour Abuse Prevention Officers, and enabled such officers to use police-level powers, as laid out in the Police and Criminal Evidence Act 1984, to investigate and enforce labour market offences occurring under the NMW Act 1998, the Employment Agencies Act 1973, the Gangmasters Licensing Act 2004 or the Modern Slavery Act 2015 (Parts 1 and 2).

In December 2018, the government introduced legislation on workplace reforms, responding to the Taylor Review into modern working practices. Although the proposed legislation is not yet public, it is likely to include substantial changes to the operating environment of the bodies, particularly for the EAS.

**Enforcement powers**
The bodies have a range of enforcement powers, ranging from approaches encouraging self-correction to more coercive options, such as enforcement notices; civil proceedings; penalty notices; LME Undertakings and Orders; licence revocation; levying employment tribunal fines (EAS); prohibition and prosecution. There is potential for their enforcement powers to expand: for example, powers to impose civil penalties as an alternative to prosecution.

**Ministerial direction**
Ministerial priorities for the EAS and NMW teams are set by BEIS Ministers. Priorities for the NMW team are also informed through enforcement data and risk modelling. Home Office ministers may direct the GLAA to prioritise areas of activity.

**IT systems**
As mentioned earlier, EAS is receiving additional funding to improve its IT infrastructure by procuring a case management system. GLAA has a substantial intelligence capacity and is in the process of being connected to the police national database, meaning it will share intelligence with and be able to access intelligence from all police forces across the UK. The NMW team has access to HMRC's risk modelling tools that utilise the range of data available to HMRC to provide analytical insight into NMW risks across the UK labour market.

**Links between LME bodies**
Partnership working between the three LME agencies has been steadily increasing. The Director's Strategic Co-ordination Group (SCG) was established by the Director of

---

[4] The Department for the Economy in Belfast is responsible for enforcing employment agency legislation in Northern Ireland.

LME to provide a strategic operational enforcement link between the Director's enforcement strategy and the three enforcement bodies.

**Wider governmental and non-governmental links**
All three agencies work in collaboration with other government departments and agencies and other organisations to carry out their intelligence-gathering and investigative work.

EAS works with the Pensions Regulator, the Health and Safety Executive, Home Office Immigration Enforcement, the Scottish Government, individual police forces, SAFERjobs[5], Department for the Economy (Northern Ireland), trade associations and trade unions.

The NMW team works and consults closely with BEIS and a wide range of different groups and stakeholders, such as the Labour Market Director, Low Pay Commission, trade unions and employer representatives to develop their understanding of their customers and issues affecting compliance with the law.

The GLAA works with partner agencies in about one third of investigations, mostly with police forces, and occasionally with local authorities, the National Crime Agency, the HSE, HMRC, Immigration Enforcement, the SIA and UK Borders.

## 2.1.3 Activities

## Awareness raising

The EAS has engaged in a range of activities to raise business and worker awareness of its role in order to support better compliance and worker consciousness. Activities include guidance on employment agency issues on GOV.UK; pop-up stalls in key businesses, the recent launch of their website and targeted campaigns to raise or refresh awareness among local authorities; ACAS advisors; and partners in the Romanian and Bulgarian Embassies to improve communication with agency workers planning to come to the UK.

The NMW Promote team encourages compliance through communications with employers and workers to raise employers' awareness of their obligations and workers' understanding of their rights. Examples of their communications activities include digital education products, such as webinars, sectoral guides, guides and guidance on gov.uk, a self-serve NMW calculator, fact sheets, leaflets, online forums and large-scale email and text campaigns. In addition, they have supported the design of commercial compliance toolkits. The NMW team has also put in place a voluntary declaration facility to encourage employers to self-review their payroll systems and self-correct minimum wage underpayments without getting HMRC involved.

GLAA actively seeks engagement with businesses and worker organisations in targeted and licensed sectors to raise awareness of how to identify potential forced labour. Last year, it ran approximately 140 training events to meet this aim. The GLAA also aims to develop industry-specific prevention officers whose role it will be to raise awareness in those industries. Like EAS, the GLAA works with embassies of countries from which victims of forced labour often come, such as Romania, to help raise awareness in those countries around what to expect from job offers in the UK, as well

---

[5] Through SAFERJobs EAS works with industry partners, law enforcement agencies, and the Department for Work and Pensions to identify and remove fraudulent online job adverts or those linked to non-compliant employers and other forms of labour market abuse.

as engaging with migrant faith groups in the UK to spread information about labour exploitation to congregations.

## Intelligence gathering

For the NMW team, intelligence comes through nationally and internationally to the HMRC risk intelligence bureau and relevant information is then passed on to the NMW team. The team also has an online complaint form and directly engages with the taxes teams in HMRC who highlight the risk of NMW infringements. Intelligence about NMW risks and unpaid positions also comes through a wide range of partners e.g. police, Home Office, DWP as well as trade unions and lobby groups. Intelligence goes through a triage team to assess whether it should get investigated.

GLAA intelligence is drawn from investigations, industry actors, cooperation with other agencies, and the public. It is now connected to the Police National Database and can access intelligence from all UK police forces. Its industry-specific prevention officers help to generate intelligence as well as raising awareness.

For EAS, intelligence comes through an increasing number of complaints, and through intelligence-led working which is informed by risk indicators and partnership with enforcement bodies and key stakeholders to identify non-complaint behaviours.

## Stakeholder engagement and co-production

The NMW team assist businesses in understanding how the law on minimum wage works in practice by producing user-friendly sector-specific guidance, online forums to answer employer queries and by supporting the design of commercial compliance toolkits.

The GLAA engages with industry stakeholders in high-risk sectors to co-design tailored guidance, protocols and codes of conduct. The aim is to ensure understanding of regulations across stakeholders, agree signs of non-compliance and decide on procedures for promoting compliance.

## 2.1.4 Outputs

There are three main outputs that result from these activities. Firstly, through promotional and educational work, the bodies seek to **increase awareness** among the public about signs of exploitation, among workers of their rights and how to seek help and among employers of their liabilities and of LME agencies' capabilities. Secondly, as a result of proactive and reactive intelligence gathering LME bodies conduct **investigations or targeted enforcement activities** and can use the range of enforcement powers described above to achieve compliance. Unless the labour market infringements are very serious, complex or large scale in nature, the enforcement powers utilised begin on the less severe end, encouraging employers or agencies to self-correct in the first instance, before increasingly more stringent enforcement activities are imposed. Finally, as a consequence of stakeholder engagement and co-production the bodies produce **agreed guidance and protocols** specific to particular sectors of the labour market and related industries.

## 2.1.5 Outcomes

The primary outcome the bodies seek to achieve is **voluntary compliance** amongst those they have not directly investigated through their awareness-raising work and stakeholder engagement, as well as through a 'ripple effect' of investigations. Linked to this, a key outcome is an **improved understanding of labour market regulations**

among employers of their liabilities and of LME agencies' capabilities; among workers of their rights and how to seek help; and among the public about signs of exploitation. Improved understanding among these groups through awareness-raising is also intended to prompt **improvement in the quality and quantity of intelligence,** resulting in an increase in more substantiated complaints and other forms of intelligence.

A key outcome is to **cease unlawful practices** among those investigated, but direct action of this kind, particularly involving the use of enforcement powers, is seen as a last resort. Through investigation and uncovering unlawful practices, a final but no-less important outcome for the LME agencies is to **recompense employees**; in the form of NMW arrears, unpaid wages or unpaid Employment Tribunal awards.

| Table 2:1 | Summary of the enforcement bodies[6] | | | | |
|---|---|---|---|---|---|
| **Body** | **Funding (£m)** | **2017/18 FTE staff** | **Focus, scope and key sectors covered** | **Locations covered** | **2016/17 cases** |
| NMW | 2017/18: 25.3 2016/17: 20 | 399+ | Covers all sectors. All employers and workers in scope, covering around 2m workers in low-paid jobs | UK | 2,674 closed 2,775 opened |
| GLAA | 2017/18: 7.1 2016/17: 4.8 | 104 | Licences labour providers in agriculture, horticulture, shellfish gathering, food processing and packaging. Over 1,000 licensed labour providers, supplying around 0.5m workers. Modern Slavery: estimated 10-13,000 victims | England, Scotland, Wales and by order in Northern Ireland | 247 licensing cases |
| EAS | 2017/18: 0.75 2016/17: 0.5 | 12 | Covers employment agencies 18,000 Employment Agencies, covering 1.1m workers | England, Wales, Scotland | 142 targeted inspections 750 complaints cleared 828 complaints received |

## 2.2  Barriers and enablers to impact

The in-depth interviews with key informants from each LME body conducted as part of this scoping study offered insights into barriers and enablers for the success of the agencies' work. These barriers and enablers are grouped below by those relating to the LME bodies themselves; those related to employers; and those relating to workers and the general public.

---

[6] Source: Metcalf (2018) United Kingdom Labour Market Enforcement Strategy 2018/19.

## Enablers within the bodies:

- **Increased level of staffing:** all three bodies have or plan to increase staff numbers to undertake a range of compliance tasks, from encouraging voluntary compliance to investigatory and enforcement actions

- **Enforcement powers:** the range of powers available to the agencies were viewed favourably. Combining different forms of enforcement where necessary was noted as effective, as was having flexibility in the degree of coercion versus cooperation

- **Joint working and triage:** cooperation between the LME bodies and partner agencies mean that investigations are quickly assigned to the most appropriate lead, depending on the nature of the suspected breach

- **Growing intelligence:** the intelligence available to the bodies is growing steadily, particularly through closer cooperation between LME bodies and with other agencies, as well as expanding LME work into new sectors. In addition, complaints have increased due to greater public awareness of workers' rights and improvements to reporting tools such as the pay and work rights online complaint form

- **Ability to deal with the higher end of the market:** the NMW team noted that they now have the capability to deal with the top end of the market, which includes companies with more complex structures, and believe efforts are being made by these companies to get their affairs in order.

## Barriers within the bodies

- **Staff resources:** despite recent and planned expansion in staff numbers, the workload of potential investigations continues to pose a challenge particularly considering increasing intelligence

- **Enforcement options:** certain enforcement options, specifically financial penalties under RESA 2008, have been discouraged by politicians, which hampers the flexibility of the bodies

- **Prosecutions:** can sometimes have poor outcomes, especially in terms of low fines, compared to the time and effort required to bring cases

- **Evidence on deterrence:** there is a need for further evidence of the 'best mix' of enforcement options to encourage compliance.

## Enablers among employers

- **Appetite for compliance:** it was thought that in most cases employers wish to be compliant and that for many the barrier is a lack of knowledge and understanding of requirements rather than intentional misconduct. Reflecting this, the GLAA commented that some businesses in licensed fields are thought to see having a licence as a mark of their integrity; and that even in unlicensed fields, employers may seek some form of licensing to demonstrate compliance. Stakeholders from the NMW team noted that the voluntary compliance portal has been well used.

- **Ripple effect:** Investigations are said to prompt other businesses in the sector to order their affairs, especially if outcomes are written about in the trade press and in public.

## Barriers among employers

- **Knowledge and understanding:** some employers, especially smaller employers without dedicated human resources staff, find it difficult to keep abreast of requirements or understand the complexities of labour market rules and regulations

- **Costing-in the price of getting caught:** some employers are thought to take account of the likelihood of getting caught, and the financial and other implications of this, in making business decisions and are therefore undeterred from non-compliance by the available penalties

- **Phoenix companies:** these companies deliberately liquidate to avoid complying with legislation or to avoid paying fines and set up in a new name, transferring their assets to the new company. Essentially, they continue the business as a new legal entity.

## Enablers among employees and the public

- **Increased media attention:** due to increased media attention, recent years have seen an increase in public awareness relating to workers' rights and public opinion is generally in favour of fair and equitable treatment of workers. This is thought to have led to more complaints from workers and the wider public.

## Barriers among employees and the public

- **Poor quality and non-actionable intelligence:** intelligence from the public can however be of poor quality or may not in reality reflect exploitation. This is partly due to the difficulty in assessing whether labour exploitation is truly taking place, and partly because the signs of exploitation may in fact point to other issues (e.g. in the GLAA Spot the Signs leaflet).

- **Fear of reprisals:** after instigating complaints employees can stop cooperating with investigations or prosecutions and withdraw complaints against their employers due to the fear of reprisals such as dismissal, losing hours or otherwise being seen to 'cause trouble' at work. Migrant workers may also fear deportation as a result of engagement with state bodies.

# 3 Literature review

One of the most difficult issues to address in any evaluation is measuring impact. Progress towards targets in most of the other stages of the evaluation logic model – such as inputs, activities and outputs – can be easily measured using administrative data that is routinely collected as part of the agency or department's ongoing work. Even some outcomes, such as increases in the number of employees recompensed as part of compliance efforts, can be recorded or at the very least inferred from data that is collected in the process of the bodies' statutory work. However, measuring impact, the final stage in the logic model, necessitates at the very least the measurement of levels of compliance or non-compliance before and after a period of activity or between comparable settings. Even then the issue of attributing causality is not entirely straightforward.

The measurement of impact in the case of the work of the bodies is likely to be particularly difficult in that the overall measure of interest, compliance or non-compliance with labour regulations, is by definition hidden. Therefore, two brief reviews of literature were carried out in order to:

1. Understand the approaches applied in the UK and more widely to evaluate the impact of labour market enforcement
2. Explore what methodologies to the measurement of compliance/non-compliance have been adopted elsewhere.

This chapter looks at each objective in turn to identify possible methodological approaches that could be adopted for use in any evaluation.

## 3.1 Approaches applied in the UK and more widely to evaluate the impact of labour market enforcement

There is relatively little evidence specifically about the impact of labour market enforcement, although many of the studies commissioned by the Low Pay Commission have assessed the quantitative impact of the rate at which the minimum wage is set (e.g. Dolton, et al., 2015, Fidrmuc & Tena, 2017). Reflecting the small evidence base, a "light touch" literature review was conducted to understand the approaches applied in the UK and more widely to evaluate the impact of labour market enforcement. It considered the following research questions:

- Is there any UK or international evidence of successful impact evaluation of labour enforcement bodies?

- Which impact evaluation approaches have been implemented? Which variables and/or proxies were used most effectively?

- Was the chosen approach robustly implemented? How was this achieved?

- Can the approach used make a causal claim? If applicable, could this be robustly extrapolated to a national level?

- What, if any, were the barriers and facilitators to successful implementation and how were these overcome?

This chapter sets out the inclusion and exclusion criteria of the review, before exploring the studies found. Whilst there are numerous studies, in the UK and abroad, that have assessed the impact of the minimum wage on outcomes such as employment, hours and earnings, this review is specifically interested in labour market enforcement.

Therefore, unless the methodology could be applied to labour market enforcement, these studies are not explored in this report.

The studies have been grouped based on their methodology. Initially, the studies using an impact evaluation approach are explored. This is exclusively limited to quasi-experimental designs, as the review did not reveal the use of randomised controlled trials in labour market enforcement evaluations. The review then goes on to explore other quantitative studies that could be relevant for future research on labour market enforcement.

### 3.1.1 Inclusion and exclusion criteria

The expected relative scarcity of studies on labour market enforcement necessitated using a very limited set of inclusion and exclusion criteria.

Inclusion criteria:

- Studies which use randomisation (i.e. a randomised controlled trial) or a quasi-experimental design (QED) that could be suitable for a causal claim. This could include: Regression Discontinuity Design (RDD), Difference-in-Differences (DID); Propensity Score Matching (PSM), Synthetic Control Groups or an Instrumental variable approach[7]

- Studies which use quantitative approaches but do not employ randomisation or a quasi-experimental design, such as longitudinal analysis or other quantitative survey analysis

- Peer-reviewed journal or "grey literature" from a trusted source (e.g. a government department or similar trusted body, such as the Low Pay Commission).

Exclusion criteria:

- Qualitative studies

- Other grey literature.

A full list of the search terms used in the literature review can be found in Appendix E.

### 3.1.2 Studies using randomisation or quasi-experimental design

The use of randomised experiments in social science is relatively uncommon, for several practical and ethical reasons. Particularly within a labour market enforcement context, the ethics of randomising individuals to receive a new programme or not could be controversial. Consequently, whilst randomised controlled trials have been growing in popularity in the last decade in the UK, particularly within education, they have not been routinely implemented in a labour market context.

The difficulty of implementing randomised experiments meant that the literature review did not find existing evidence of RCTs being used to evaluate labour market enforcement programmes, either in the UK or abroad. However, several studies included in the review used quasi-experimental design approaches.

---

[7] Table 1:1 explains these different approaches, with examples. Alternatively, readers can consult "The Green Book" (HM Treasury, 2018).

## Instrumental Variables

One of the most commonly used approaches identified by the review was the use of instrumental variables (IV). A U.S. study (Galvin, 2016) investigated the impact of the strength of labour market regulations on minimum wage violations using a dataset created by researchers that ranks U.S. states on the strength of the labour market regulation over time. In the U.S. there is relatively weak labour market regulation at the federal level, but a wide range of regulations implemented at a state level. Using the IV approach, the authors found evidence of stronger state regulation reducing the number of minimum wage violations. They then extended the analysis, comparing changes in state regulation over time using a difference-in-differences (DID) approach, which reinforced their original finding.

Although Galvin's approach is robust, it is not one which is applicable to the UK context. A key element of the study is that there is enough variability in the strength of labour regulations across states that these states can be compared. By contrast, labour market regulations in the UK are very similar across countries. Whilst there are some regulations that only apply in England, or England and Wales, this is not enough to estimate variability at the country level.

The IV approach was also adopted in two similar Argentine studies of labour market enforcement. Ronconi (2010) instruments the strength of labour market enforcement bodies using the per capita number of inspectors, whilst Viollaz et al. (2018) take a more complex instrument to estimate the "arrival cost" of inspectors. The approach overall is similar in both studies. Viollaz et al. is explored in more detail below, as it is both more recent and the chosen instrument more complex.

The first stage equation instruments the strength of enforcement by instrumenting the "arrival cost" of inspectors, using data on the share of inspected firms by province and sector over time. To instrument the arrival cost, the author focuses on transport costs for inspectors, using expansion of the road network and traffic congestion. The first stage equation also includes productivity, as inspectors tend to target firms with expansionary activities. The second stage model estimates the compliance with social security regulations, controlling from the instrument and import individual level characteristics.

This analysis was conducted separately for men and for women to explore if there is a differential impact of enforcement for men and women. Overall, the author found that increased labour enforcement increases compliance for men, but women experience a decrease in compliance, with both formal and informal wages decreasing.

Although the approach is robustly implemented, there are several key limitations to this study. Firstly, the study relies on household survey data. Whilst the coverage of the EPH (translated to Permanent Household Survey) is good for urban areas (70% of the urban population, Ministerio de Economia, Infraestructura y Energia, 2019) it does not cover rural employees. Whilst coverage of the urban and rural population is not an issue for UK surveys, coverage is still a significant issue to consider. Particularly relevant to evaluating non-compliance with labour regulations is the known poor survey coverage of households at either end of the income distribution that are known to be under-represented.

Finding a convincing instrument for labour market enforcement would also be particularly challenging, as issues such as road network expansion are unlikely to be suitable in the UK context.

## Difference-in-differences

Although some studies combine DID with other approaches (e.g. Galvin (2016) which combined IV and DID) the methodology is suitable to be applied in isolation. An example of this from South Africa employed DID to estimate the causal impact of introducing minimum wages on non-compliance with labour regulations (Bhorat, et al., 2015).

In South Africa, there is no national minimum wage, but there are some sectoral minimum wages. The focus of this study is the introduction of a minimum wage in agriculture in 2003, which was introduced with slightly higher wages in urban areas than rural areas. To account for changes in the overall economy, the researchers employ DID, comparing the change in the difference between those subject to the minimum wage and workers with similar characteristics that were not subject to any minimum wage legislation[8].

They then analyse the gap between reported wages (from the South African Labour Force Survey) and the minimum wages (extending the definition of minimum wages to those in the control group). They found evidence that introducing the minimum wage law led not only to an increase in wages, but substantially reduced the gap in underpayment.

Again, the use of survey data raises questions about coverage, and it is important to note that whilst the authors applied strict criteria to which workers are in the control group, they do not combine DID with other methods such as propensity score matching which could reduce bias between groups. The approach may be applicable to the UK context, if suitable data can be found to measure non-compliance (either with minimum wages or other labour market regulations).

## Regression Discontinuity Design

Gindling, et al. (2015) explore the impact of a concerted publicity campaign alongisde increased enforcement to assess the impact on payment of minimum wages in Costa Rica. Compliance with minimum wage legislation is known to be low (potentially as low as 30%) and the context is therefore different to that of the UK. The campaign involved three distinct elements:

- A publicity campaign to raise awareness among employers and workers regarding the importance of minimum wage compliance

- Encourage employees to report employers who pay less than the minimum wage, including a new complaints phone line

- Increased labor inspections, targeting non-compliance with the minimum wage

Using household survey data (Costa Rican National Household Survey) the authors analysed the change in wages for two groups, 10% above and 10% below the minimum wage, between two periods before and after the campaign was introduced. They found that the campaign increased compliance with the minimum wage.

This study uses a robust methodology that has also been used in the UK in similar contexts. For example Fidrmuc & Tena (2017) used RDD to assess the impact of progressing from NMW youth rates to the adult rate on the likelihood of being in

---

[8] The comparison group is made up of employees in unskilled or 'elementary' occupations, based on the 4-digit SASCO occupation codes and ISIC industry codes, earning less than the Basic Condition of Employment Act's (BCEA) income cut-off of R9 631 per month, aged between 15–65, who have completed no more than 12 years of schooling. In addition, union members, and those in sectors affected by another minimum wage, are excluded.

employment for young people in the UK. However whilst this design can be used where an administrative cut-off decision is made, it is difficult to apply this to other contexts within labour market regulation.

### 3.1.3 Other quantitative studies

While not all the studies included in the literature review use a quasi-experimental design, some still utilise methodologies that could be implemented to improve the current evidence base on the effects of labour market enforcement in the UK.

One such study is an Australian telephone survey of employers, exploring perceptions of labour market enforcement activities (Hardy & Howe, 2017). The telephone survey explored four elements of enforcement: targeted campaigns, compliance notices, enforceable undertakings, and proactive compliance deeds. The survey was undertaken by an independent evaluator to reduce the likelihood of social desirability bias. Another Australian study (Parker & Nielsen, 2011) also explored employer perceptions of the enforcement of labour market regulations using a postal survey with telephone follow-up.

Another approach is to assess trends using administrative data sources. An example of this is a study from Canada (Vosko, et al., 2017) which used administrative data from the Employment Standards Information System (ESIS) provided by the Canadian Ministry of Labour. Although this study could not conduct counterfactual analysis, it provided useful insights into how cases of reported non-compliance were handled. However, using administrative data cannot be used to establish causal impacts as the coverage of the data is limited to only those in contact with the relevant labour market authorities and cannot be applied to those who have not reported non-compliance.

Other studies used survey data in regression analyses to assess various aspects of non-compliance, but without employing an impact evaluation technique. For example, some studies used a respondent-led sampling survey to understand variations in non-compliance between U.S. cities (Milkman, et al., 2012, Bernhardt, et al., 2013).

## 3.2 Methodologies for measuring compliance / non-compliance

Due to the need to baseline the level of non-compliance in order to undertake an evaluation of the LME bodies, a brief review of the literature was carried out to explore what methodologies for the measurement of compliance/non-compliance have been utilised in the UK and elsewhere and to identify possible methodological approaches that could be adopted for use in any evaluation. The review incorporated a targeted search of academic databases and likely sources of grey literature identified by the team, as well as recommendations provided by colleagues with knowledge in the area of labour market regulation.

A search of the literature identified, firstly, that there is a relative paucity of research that has successfully measured levels of non-compliance with the sorts of employment and labour market regulations that are within the remit of the bodies and, secondly, that most of the extant literature consists of cross-country studies investigating relative levels of non-compliance in countries where non-compliance is relatively high. The lack of relevant research in this area has been identified elsewhere. For example, Ritchie, Veliziotis, Drew and Whittard (2016) note that little research about the scale of minimum wage non-compliance in high-income countries, while Belman and Wolfson (2014) report that there are over 200 policy and academic reports published between 1992 and 2013 looking at minimum wages, a search of Google Scholar reveals that

only a handful actually look at non-compliance and many of these do not consider the accuracy of their estimates.

That said, the table in Appendix F contains a summary of some of the most relevant studies that were identified in the literature, including a brief summary of the data and methodology used to measure non-compliance as well as the key advantages, disadvantages and learning points from each study. Broadly, three main approaches were identified from the literature:

1. **Approaches using Household surveys** – often using data from the equivalent of the Labour Force Survey (LFS) in different countries to look at non-compliance with minimum wage legislation, mainly using self-reported income data.

2. **Approaches using targeted surveys** – one-off surveys used to access hard-to-reach populations and/or attempting to measure other types of non-compliance (not just compliance with minimum wages).

3. **Approaches based on administrative data and targeted enforcement activity** – often aiming to identify or measure non-compliance in industries or areas where there are particular concerns about non-compliance.

## 3.2.1 Approaches using household survey data

As can be seen from the summary table in Appendix Table F, the majority of studies in this category consist of cross-national studies that attempt to measure levels of non-compliance (usually with minimum wage legislation) in different countries in order to assess the extent to which non-compliance may be driven by different factors (such as regulatory frameworks or market factors). These sorts of studies exploit differences in any number of explanatory characteristics of different countries and/or regions in order to explain variation in levels of non-compliance.

Only three of the studies identified in the literature attempted to estimate levels of non-compliance within just one country, either in the UK (Ritchie et al, 2016; Low Pay Commission [LPC], 2017), or in Australia (Nelms et al, 2011). Ritchie et al compared rates of non-compliance with the minimum wage among apprentices using data from Apprenticeship Pay Survey (APS) and the Annual Survey of Hours and Earnings (ASHE), for which detailed information on income is sought from employers for a random sample of workers. The LPC used ASHE to estimate non-compliance with National Minimum Wage / National Living Wage in the UK. Using four sources of survey data Nelms et al attempted to more accurately estimate levels of non-compliance with federal minimum wage regulations in Australia.

The advantages of using household survey or other statutory survey data can be summarised as follows:

- The data is routinely captured as part of statutory requirements, so is easily accessible and inexpensive and can be monitored over time.

- While concerns have been raised in the literature about the accuracy of estimates of non-compliance using household and statutory survey data, it has been argued that the accuracy of measures of non-compliance are of secondary importance in studies where the aim is to estimate the functional relationships between different factors or variables (Ritchie et al, 2016).

- Limitations in some datasets may be offset by combining data from different datasets or triangulating findings, depending upon the objectives of the research. For example, while data from the Annual Survey of Hours and Earnings (ASHE) is often assumed to have more accurate income data the dataset does not cover

informal employment or contain many personal characteristics, whereas the LFS may be more likely to capture some informal employment and has a range of personal characteristics of individuals but is likely to suffer from some measurement error.

On the other hand, some of the disadvantages are:

- Multivariate analyses of the impact of different factors on non-compliance (including potential impact of enforcement activity) rely on being able to model for variation in the different inputs or variables in different geographies, industries, or constituencies. This may not be possible where the same regulations, activities and inputs have been applied at a national level.

- Household surveys rely on self-report data (recall, honest answers, rounding – LPC, 2017; Ritchie et al, 2016) and may miss hard-to-reach or vulnerable communities, underestimating the informal sector (Bernhardt et al, 2009).

- Many low and middle-income countries have high levels of non-compliance and so inaccuracies in estimates are less of a concern (Ritchie et al, 2016).

- Nelms et al (2011) identified a number of measurement issues (e.g. estimating hourly earnings from weekly data) and also identified five reasons for finding below minimum wage individuals, some of which do not indicate non-compliance

  1. Exclusions (e.g. training wage)

  2. Long hours with no paid overtime in reference week

  3. Does not account for non-wage benefits (e.g. pension contributions)

  4. Genuine non-compliance (e.g. payment of wages below the relevant rate or making illegal deductions)

  5. Measurement error (these are more of a problem in cases where estimated non-compliance is low – 7-9%).

- In most cases the questions covered by statutory surveys are not designed to capture non-compliance with employment regulations (other than pay) and cannot be easily redesigned, limiting the types of non-compliance that can be studied using this data.

## 3.2.2 Approaches using bespoke or targeted surveys

A number of studies identified in searches used targeted surveys focused on specific sectors or forms of non-compliance and/or employed alternative methods to sampling in order to better target hard-to-reach populations (see Appendix F). Weil (2005) surveyed employers in the Los Angeles apparel industry by randomly selecting from the Californian Manufacturers Register and applying weights based on assumptions about estimated non-compliance. Bernhardt (20015) and Frieberg (2012) used respondent driven sampling (RDS), which uses 'chain-referral sampling' but weights respondents based on size of their network (and observed biases), to survey low-wage workers and migrant workers respectively.

The benefits of these approaches can be summed up as:

- Enables the researcher to focus the questionnaire on a range of forms of non-compliance.

- Can be used to get a representative/generalisable sample of a hard-to-reach population that would otherwise be missed.

- Design effects can be estimated and validation checks carried out.

On the other hand, limitations of these kinds of approaches are:

- These types of surveys can be labour-intensive and costly to run

- Targeted surveys either rely on a sample frame being available, as in the Weil (2005) study, and surveys using a method such as RDS are only appropriate under certain conditions/assumptions

- Incentive levels need to be appropriate (not too high/low), or it may encourage false responses

- It is important to be able to communicate the criteria clearly and simply using RDS.

Drawing on several studies using RDS, Friberg and Horst (2014) discuss the appropriateness of using RDS to research a specific hard-to-reach population, namely migrants. They conclude that in many cases migrants are well suited to the assumptions of RDS, namely that: respondents know each other and recognise each other as being in the target population (have pre-existing contact patterns and that these are reciprocal); and there is sufficient crossover between subgroups in the population of interest. The authors note that the approach relies on natural social groups and the ties that they have. Successfully applying the approach may involve a process of defining and re-defining the target population as the research evolves. Scoping research and identification of bottlenecks is important to the success of the approach.

## 3.2.3 Administrative data and targeted enforcement action

The final approach identified in the literature in relation to measurement of non-compliance was based on the use of administrative data and/or targeted enforcement action. The majority of the studies outlined in this section in Appendix F either come from the work of Weil and colleagues (e.g. Weil, 2012; Kleiner and Weil, 2010; Ji and Weil, 2015) in the US or were carried out or commissioned by the Fair Work Ombudsman (FWO) in Australia.

The work of Weil and colleagues has used administrative data to explore levels of compliance with federal minimum wage and overtime standards in franchised and company-owned establishments in the fast food industry in the US; the effect of different ownership models on compliance / non-compliance with wage regulations in the fast food industry and the hotel and motel industry and the effectiveness of penalties as a deterrent to non-compliance.

The FWO in Australia routinely publishes five different types of report aimed at examining non-compliance with regulations:

1. **Annual reports** – FWO annual reports, performance over the year (outputs)

2. **Research reports** – Research to understand specific workplace issues (e.g. phoenix activity – PwC, 2012; unpaid work arrangements – Stewart and Owens, 2013)

3. **Submissions** – Reviews of FWO activities and effectiveness carried out by external researchers (e.g. Howe, Hardy and Cooney, 2014)

4. **Activity Reports** – Following allegations of serious non-compliance the FWO publish the findings of their compliance activities with specific organisations

5. **Inquiry reports** – Reports from campaigns of focused enforcement activity in specific industries, areas, regions, supply chains or labour markets, in response to reports of systematic non-compliance highlighted in official data, media stories or public concerns (e.g. harvest trail, local government services procurement, supermarket cleaners, working holiday visa workers).

The examples of FWO reports identified in Appendix F are from categories 2, 3 and 5 above. These types of approaches can, at best, only be used to infer wider levels of non-compliance or impact over time. However, adopting certain elements of these approaches could be used in combination with other approaches outlined above and could be employed where targeted action or enforcement activity is already planned. Some of the advantages could be seen as:

- Targeted enforcement activity can give a good idea of levels of non-compliance within a specific area, industry or market, although perhaps not broadly generalisable

- Data from these sorts of activities may be the only source of data on certain types of non-compliance

- Data can be used to monitor progress going forward through revisits (sustained impact)

- There is the possibility that, where inspections or audits are random, the data could be generalised more widely to similar organisations

- While data collected from enforcement activities may not be broadly generalisable, it may give indications as to relative non-compliance in different types of organisations.

Some of the limitations can be seen as:

- Because enforcement activity is normally in response to a complaint or is targeted at industries or areas where there are particular concerns, levels of non-compliance found are not normally generalisable

- Carrying out targeted enforcement activity is costly if not already budgeted for in the normal business of the enforcement body or agency

- Resources may often need to be focused towards responding to complaints about non-compliance for ethical reasons.

# 4 Evaluation design options

The original aim of this scoping exercise was to evaluate the impact of labour market enforcement, with particular regard to the three bodies[9] under the remit of the Director of Labour Market Enforcement. The study findings suggest that broad scope and complexity of labour market enforcement and the overlapping powers and responsibilities of each of these bodies makes a single evaluation encompassing all aspects of enforcement unfeasible at this time. In particular:

- The logic model identifies five distinct outcomes, which are achieved through several causal chains (i.e. the way each body intervenes is different, even if the strategies are similar)

- The optimal research design to assess the achievement of these outcomes therefore varies between outcomes

- The available data to assess these outcomes varies between each of the three enforcement bodies

- Some of these outcomes are not suitable for an impact evaluation attempting to make a causal inference.

This chapter outlines several suggestions for potential evaluation designs to provide credible options for the Directorate of Labour Market Enforcement, including estimates of duration and costs for each of the designs proposed. As a single evaluation is not currently feasible, this chapter sets out credible options for analysis that could be included in a contribution analysis. Each of the following research questions below will be explored, in relation to the three work streams detailed in the logic model[10]. The designs explored in this chapter do not cover all the outcomes illustrated in the logic model, as it focuses on where robust methodologies could be implemented.

**Public communications:**

- To what extent are "promote" activities effective at raising awareness and increasing understanding of regulations?

**Direct Action:**

- Does enforcement increase compliance with regulations amongst similar firms?

- Do the additional powers LME bodies have in England and Wales affect compliance?

- How does enforcement improve outcomes for workers?

**Prevention:**

- How effective are "prevent" activities at increasing the quality and quantity of intelligence gathered by LME bodies?

The chapter then presents three options for measuring non-compliance with labour market regulations, relevant to each of the bodies.

---

[9] GLAA, HMRC NMW team and EAS.
[10] These are: unlawful activities ceased, employees recompensed, voluntary compliance amongst those not investigated, improved understanding of labour market regulations and improved quantity and quality of intelligence.

## 4.1 Potential data sources

Even if an evaluation design could be robustly implemented in theory, data limitations often prevent some methodologies from being employed. This section explores the available data sources that could facilitate future evaluations of labour market enforcement.

## Survey Data

Data collection is time-consuming and expensive. Where possible, it is therefore helpful to explore existing data sources, including large representative surveys that are regularly conducted with a focus on the labour market. This may include the Annual Survey of Hours and Earnings (ASHE), the Labour Force Survey (LFS) and the Family Resources Survey (FRS).

**The Annual Survey of Hours and Earnings**
ASHE is the largest of these surveys, consisting of a one percent sample of National Insurance numbers (NINos). Unlike other surveys, ASHE is issued on an annual basis to employers who complete the survey on behalf of employees. Consequently, the survey only has coverage of people in employment.

As employers complete the survey, the earnings data in ASHE is actual earnings and is therefore considered to be more reliable than other surveys, such as the LFS which relies on individual self-reporting. The large sample size in ASHE also facilitates analysis of specific groups, particularly within industries and occupations.

However, the short survey length means that some characteristics, such as people's educational qualifications and household composition, are not collected, which can restrict some analyses from being undertaken with ASHE data. It is also possible that some employers may fill in the survey inaccurately if they underpay employees, due to social desirability bias. In practice a reasonable proportion of individuals in the survey (1.2% in 2017/18) appear not to be paid the minimum wage based on the information employers have provided[11] (BEIS, 2018).

**The Labour Force Survey**
The LFS is a quarterly survey of households, with approximately 40,000 respondents in each quarter. The survey design involves recruiting new households in each quarter and following them for five quarters. The strength of this survey design is that the dataset can either be analysed cross-sectionally or longitudinally.

However, as earnings data is only collected in the first and last quarters respondents are surveyed, there are some limitations on the analysis that can be conducted. The LFS does collect a much wider array of information than ASHE, such as educational qualifications and household composition, which facilitates analysis using more data-intensive techniques. Questions on the LFS are also harmonised with similar surveys conducted in other European countries, which are consolidated into the EU-LFS, which facilitates cross-country analysis.

As with all household surveys, non-response can affect coverage, particularly at the bottom and top of the income distribution, which should be considered in the context of analysis of labour market enforcement. The earnings data in LFS is self-reported and consequently earnings data in LFS is considered less reliable than in ASHE. Where

---

[11] Although it is possible that some cases may be as a result of inaccurate reporting, the proportion is large enough that this should be considered as genuine minimum wage non-compliance.

possible, interviewers are instructed to check respondents' payslips, if they are available, to validate the earnings reported in the survey.

**The Family Resources Survey**
The FRS is an annual survey of households incorporating a wide range of questions on family's financial resources, including extensive questions on housing costs and assets. The FRS is also used by the Department for Work and Pensions (DWP) to create the Households Below Average Income (HBAI) dataset. The key strength of the Family Resources Survey is the wide range of topics it covers, allowing researchers to assess a number of topics, such as housing costs, assets and deprivation.

As with all household surveys, the FRS is known to suffer from coverage issues, both at the top and bottom of the income distribution. It also includes accurate information on benefit receipt, as this is information is added to the FRS data directly by DWP. The FRS is also a little smaller than the LFS, covering approximately 20,000 households. Whilst it could be used for analysis of labour market enforcement, the benefits of a large array of characteristics are outweighed by the limitations of the smaller sample size and therefore ASHE and the LFS may be more appropriate for analysis in this context.

## Data collected by enforcement bodies

Each of the three enforcement bodies collects data pertaining to the individuals and employers in their remit. In many cases, this information is already analysed by each body, providing useful indicators of their activities. However, most of the data collected by the bodies relates to individuals who have raised a complaint, or employers that have been inspected or investigated. The coverage of the data therefore limits the quantitative analysis that can be undertaken, as it is impossible to establish a counterfactual if the coverage of data only includes people who are in direct contact with the LME bodies.

**HM Revenue and Customs**
The HMRC NMW team have access to the most data out of the three enforcement bodies, a virtue of HMRC's wider remit and resources. The NMW team strategy focuses on three elements: promotion, response and targeted enforcement.

The promote function of HMRC often focuses on areas of "accidental non-compliance" amongst employers, with the aim of increasing voluntary compliance. Administrative data collected as part of the promote function gives a quantitative indication of reach. In 2017/18, HMRC's 'promote' function enabled the following contact[12]:

- 11,700 webinar attendees, with high recorded satisfaction amongst participants

- 617,744 email nudges to employers

- 1.6m text messages sent in May and June to at risk workers, 1.3m to tax credit recipients and 370,000 texts to apprentices

- 56 voluntary declarations associated with £246,000 in arrears repaid to 689 workers.

The response function aims to resolve 100% of complaints raised to HMRC. In 2017/18, 6,027 complaints were raised with HMRC. Of these, 4,161 raised a complaint online, 1,790 through ACAS and 76 from other sources. The volume of complaints over time is monitored by HMRC alongside the volume of wage arrears and number of workers affected.

---

[12] Department for Business, Energy and Industrial Strategy (2018).

Finally, the targeted enforcement function is based on identified risk of non-compliance. The decision on which areas of the labour market that should be subject to targeted enforcement is determined by HMRC's internal risk modelling, BEIS, Director of Labour Market Enforcement analysis, third party intelligence and ministerial priorities.

Approximately 15-20 projects are undertaken in the targeted enforcement function each year, of which approximately 10-12 are included based on evidence from the risk model. The risk model incorporates real-time data from tax collection, ASHE data amongst a variety of other measures. All areas identified as high risk from the modelling process are included in the projects enforced. However, many areas are identified as medium risk. The risking team therefore decides which of these areas to enforce, based on intelligence and the priorities of the DLME, other enforcement bodies and ministers.

**Employment Agency Standards**

The EAS strategy is similar to HMRC's in many ways. Like HMRC, there is a focus on awareness-raising, complaint-driven activities and targeted enforcement. EAS' awareness-raising includes promotion to employment agencies, pop-up stalls in key businesses, online guidance and the recent launch of their website.

Complaints have been increasing year on year. In 2015/16, 781 complaints were raised with EAS, increasing to 828 in 2017/18 (Department for Business, Energy and Industrial Strategy, 2017). Complaints have substantially increased again in 2018/19. To avoid discouraging potential complainants, EAS do not mandate the information a complainant has to provide. This is for operational reasons but limits potential quantitative analysis.

When inspectors investigate an agency, either in response to a complaint or as part of their targeted enforcement activities, they regularly collect data in their record of inspection, which is prompted by an aide memoir to cover all the aspects of regulations within EAS' remit. However, the record of inspection only covers infringements, or details about previous infringements that are no longer an issue. This covers many aspects of the regulations in EAS' remit, such as deductions from pay, the size of the company, whether contracts are temporary or permanent and whether workers have been given a health and safety briefing.

**Gangmasters and Labour Abuse Authority**

The GLAA collects data on several aspects of enforcement within their remit, primarily related to licensing and intelligence.

Licensing covers firms supplying workers for roles in several sectors: agriculture, horticulture, shellfish gathering and any associated processing and packaging. Intelligence comes from a wide array of sources, such as complaints, the police and other LME bodies. The data collected by GLAA includes:

- **Licensing data:** which includes information on business details, personal details, contact information and details of the licence record

- **Licence decision data:** which covers applications submitted, licences granted, whether the licence has been granted with conditions, whether a licence has been refused or the licensing decision appealed

- **Intelligence:** a large amount of information, covering reports of licence breaches, possible criminality in terms of modern slavery, information on victims, third parties and sector

- **Performance data:** Cases considered, cases tasked for investigation, use of intelligence and results.

Intelligence cannot be used in future evaluations, as sharing this data with third parties is not desirable. However, licensing data, licensing decision data and performance data could all be potentially utilised by an evaluator.

The GLAA has the lowest volume of cases of the three bodies, though the breaches they investigate can be more severe, such as modern slavery. In 2018, the GLAA handles approximately 500 cases, of which roughly 200 were related to licensing activity and 300 associated with wider labour market enforcement.

Similar to the other LME bodies, the GLAA also engages in prevention activities with employers. In 2018, this included approximately 140 training events with roughly 14,500 people in businesses and other organisations.

## Future data collection

In some cases where data is unavailable, or the coverage of existing data is not sufficient for an evaluation design, it may be possible to collect data. This would usually be undertaken via a survey. In the context of labour market enforcement, this could involve surveying employers or households (i.e. workers).

Aside from cost and the time it takes to collect data, collecting new survey data is also potentially subject to issues of coverage. As noted earlier in this section, survey data tends to have poor coverage of households at the bottom of the income distribution, due to non-response. In the context of labour market enforcement, hard-to-reach groups, who may be the most vulnerable to labour market exploitation, may be less likely to respond to surveys. It is therefore important to consider how this can be mitigated in the survey design. For example, oversampling at risk groups and weighting to be representative of the general population, or by implementing a respondent-driven sampling survey, as described in section 4.5.

# 4.2   Public communications

The logic model indicates that improving the understanding of labour market regulations is a key outcome of the work undertaken by each of the three enforcement bodies. The primary research question for public communications is to understand to what extent "promote" activities are effective at raising awareness and increasing understanding of regulations, the labour market bodies and their respective complaints routes. It will be important for any future research to distinguish between the awareness and understanding of employers and of individuals.

Unlike the Costa Rican publicity campaign (Gindling, et al., 2015) explored in the literature review, the promotional activities undertaken in the UK have been a significant element for some time. Although some changes have been introduced over time, such as the "naming and shaming" of employers who fail to pay the minimum wage, there is no significant time point at which promotional activities were introduced that would lend itself to an over-time comparison of public and employer awareness. This is also exacerbated by a lack of representative data that could be analysed by an independent research body. These two issues make an estimation of causal impact inappropriate at present.

However, some administrative data already exists that could provide some insights about public awareness. Analysing the volume of contact with advisory bodies, such as the Advisory, Conciliation and Arbitration Service (Acas) over time could provide some

insights. This cannot be used to establish a causal claim as it would be impossible to eliminate potential confounders. For example, if the volume of complaints increased from one time period to the next, it could be because there is greater awareness of workers' rights, but it could also indicate a higher incidence of non-compliance amongst employers.

There are two approaches that are recommended to assess the impact of promotion activities. Randomisation of promotional materials, and quantitative surveys to robustly estimate employer and individual awareness of labour market regulations.

## 4.2.1 Randomisation to assess the impact of promotion activities on voluntary compliance

Previous randomised controlled trials (including one with up to 670,000 employers) found that the use of email "nudges" helped increase voluntary compliance with labour market regulations[13]. This could be applicable to all three labour market enforcement bodies, though the sample that is randomised would vary depending on the body being evaluated[14].

A random sample of employers, or all businesses within a specific sector, could be selected[15]. It is suggested that this would be the Inter-Departmental Business Register (IDBR) that is held by ONS. The Office for the Director of Labour Market Enforcement, or the Department for Business, Energy and Industrial Strategy, would need to facilitate access to this dataset on behalf of the evaluator. Alternatively, a commercial dataset of businesses could be purchased, but this would incur additional costs.

Within this sample, an independent evaluator could randomly allocate[16] businesses to receive a promotional activity (the intervention group) or "business as usual" (the control group). Although the previous RCT assessed the impact of email nudges, different types of promotional activities could be evaluated. These activities could include:

- **Nudges sent via email, post or text signposting information:** This would involve sending an email, letter or text(s), or a combination of the three to employers in the intervention group, signposting them towards resources, such as online guidance of a specific labour market regulation. The control group could still access the information but would not receive the nudge[17]

- **Being invited to take part in webinars or training**: The intervention group could receive an invitation to participate in a webinar or training to help them understand a specific issue (such as deduction of childcare costs from pay). The control group would not be prevented from accessing the webinars, but they would not be promoted to this group

---

[13] The original reference of this study could not be located, but is indirectly referenced in the BEIS NMW Strategy (Department for Business, Energy and Industrial Strategy, 2018) and in a recent report from the Low Pay Commission (Low Pay Commission, 2017).
[14] As the bodies have powers over different regulations, it is sensible to analyse bodies separately in this scenario because the content of the "nudge" would be sent directly from an individual body in practice. If the bodies are merged in the future, then it still may be advisable to conduct separate trials for nudges aimed at encouraging voluntary compliance in different areas.
[15] Although all business could be considered in practice, the cost of the trial would increase the larger the sample of businesses randomised.
[16] This could be implemented using randomisation, or adaptive minimisation.
[17] In a variant of this trial, the information could be sent directly to those in the intervention group.

- **Nudges sent via email, post or texts, promoting "self-correction" to employers:** If awareness of self-correction is low, the impact of a promotional activity could be randomised in a similar approach to those described above. Those in the intervention group would receive information about the possibility for self-correction. Those in the control group would still be able to voluntarily self-correct but would not receive the promotional information about it.

This approach could work to assess the impact of promotional activities but would only be possible if the administrative data on those who voluntarily self-correct could be linked with the randomisation data. Measurement may also be difficult if employers self-correct internally, but do not register this self-correction with the appropriate LME body.

It may therefore be preferable to collect outcomes data using an independent survey (i.e. data collection) but this would add to the cost of the evaluation. This is particularly true if the prevalence of the outcome is low, as more statistical power would be required to detect an effect. For this reason, it may be advantageous to run the trial in a sector where non-compliance is anticipated to be higher than average. There are risks that non-response to the survey could be higher amongst non-compliant firms and that respondents could present an overly favourable view of compliance (social desirability bias). These findings could also be compared with the administrative data described above.

In addition, the associated cost with collecting such information would restrict the size of the trial that could be implemented. Relying on purely administrative data would mean that there is little additional cost to an evaluator of increasing the number of businesses in the scope of the trial. However, data collection would only be feasible with a sample of approximately 2,500 to 5,000 businesses.

Randomised controlled trials are also expensive and take a significant amount of time to run. They must also be implemented prospectively, as the evaluator must randomise the groups to intervention or control. Ideally, the evaluator would have access to the characteristics of the businesses, to assess baseline balance between the intervention and control group. These characteristics could include sector, firm size and whether they have previously been found to be in breach of labour market regulations. If it is possible to track receipt of promotional activities (if via email or text), then this data would also be useful for an evaluator to understand the impact of dosage (i.e. the volume of emails or texts associated with impact).

Although justifying the ethics of randomisation can be difficult in some contexts, as the randomisation in this case does not prevent employers in the control group accessing the information, it is an approach that could be reasonably justified.

## 4.2.2 Survey of employers

Assessing employer awareness and understanding of labour market regulations could be robustly estimated using a telephone survey of employers, similar to the Australian studies assessing employer perceptions of labour market enforcement outlined in Chapter 3. The purpose of an employer survey would be to gauge how promotion work from the enforcement bodies affects their understanding and awareness of the regulations that apply to them.

The survey could include questions about which activities they have been involved with, such as taking part in a webinar or consulting online guidance. These could be used to inform the DLME about which types of employers are least in contact with the bodies. This could also be complemented by some simple analysis of the administrative data collected by bodies as part of their activities.

The survey could also include questions tailored to specifc priority areas for the DLME. For instance, recent promotional activities by HMRC targeted issues with deducting childcare costs directly from payslips, rather than being charged seperately. If the survey were repeated over time, it would be possible to tailor questions to these specific issues to gauge employer understanding before and after such activities took place.

The recommended approach would involve a random probability telephone survey of businesses, stratified by region, sector and firm size. The survey should ideally oversample large businesses, as the distribution of employees within firms is heavily skewed. The data would then be weighted for representative analysis.

The survey sample could be sourced from the Inter-Departmental Business Register, which is held by the Office for National Statistics. As previously mentioned, the Office for the Director of Labour Market Enfocment, or BEIS would have to facilitate access to this data on behalf of the evaluator. This sample contains a list of businesses (excluding self-employed individuals) that could be used as the sampling frame. Alternatively, the sample frame could be bought commercially, though this would incur additional cost. Once the sample is selected, the commissioned organisation would then need to find the relevant person within the business to speak to, which is often a challenging element of business surveys.

The sample size depends on the depth of analysis required. For a relatively straight-forward analysis, comparing perceptions amongst sub-groups such as gender and age, a sample of approximately 2,000 businesses should be sufficient. However, the sample size would need to be larger if analysis of smaller groups, for instance by sector, were required and the costs of implementing the survey would therefore increase. A response rate of approximately 20 to 25 percent could be realistically expected for a business survey of this type.

It would be possible to include a reasonable amount of items in the survey with the recommended length of 10–15 minutes and if required, different questions could be routed to employers from different sectors (e.g. additional questions about licensing could be referred to businesses in sectors under the juristiction of the GLAA). The survey could be longer (e.g. 20 minutes) but this will make the survey more expensive and could increase non-response.

The survey could be repeated over time to assess how understanding and awareness of employers changes over time. If the questionnaire remained the same, it may reduce the costs of repeating a survey, but it is possible that the questionnaire may need to be adapted over time as labour market regulations change. It may also be possible to incorporate some qualitative depth-interviews into this research, which would incur additional cost.

### 4.2.3 Survey of individuals

A quantitative survey could be used to investigate individuals' understanding of their labour market rights. It is recommended that a random-probability omnibus survey is used, such as the ONS' Opinion and Lifestyle Survey or the NatCen Panel. A random-probability panel is the most robust approach for obtaining representative estimates for the population and is more cost-effective than commissioning a stand-alone survey when survey length is short.

The sample for these surveys are typically drawn from existing random-probability surveys. The Opinion and Lifestyle Survey recruits participants from the LFS whilst the

NatCen Panel recruits from British Social Attitudes (BSA). The sample size on random-probability omnibus surveys are typically 1,000 to 2,500 individuals. If the survey was only concerning people in work, the sample would likely be smaller. A sample of this size should allow some basic quantitative analysis, exploring differences between sub-groups of the population.

The survey is likely to be mixed-mode, using web and phone, as this is the most efficient way of achieving robust results for a survey of this type. A survey length of approximately ten minutes would be appropriate, though it could be longer if additional items were required on the questionnaire.

An alternative approach would be to include questions on other existing market research omnibus surveys. These studies are weighted to the characteristics of the population, but they could suffer from selection bias as they do not adopt a random-probability survey design. Additionally, survey length would likely be shorter, at around five minutes, limiting which items could be included.

As with the telephone survey of employers, this survey could be repeated over time to assess how awareness of labour market rights changes over time. As before, if the questionnaire remains the same, it may reduce the costs of repeating the survey. However, if labour market rights change, it may be necessary to undertake further questionnaire development with repeated surveys.

## 4.2.4  Qualitative research with employers and individuals

Qualitative research using in-depth interviews with employers and individuals could complement the quantitative approaches described above. They would add further detail and nuance on employers' and individuals' knowledge and understanding of labour market regulations and rights, and offer insights into the sources of their knowledge and understanding.

As such, in-depth interviews, as opposed to qualitative techniques (such as discussion groups or deliberative workshops), would be most appropriate because they would allow for deeper insights and probing. It would be more challenging to go beyond surface-level insights with group-based techniques.

Telephone rather than face-to-face interviews are recommended. They offer greater flexibility, particularly for employers who can be a challenging group to recruit due to time pressures. Interviews conducted by phone can also encourage more openness about gaps in knowledge by affording employers and workers greater anonymity. Furthermore, telephone interviews are efficient and cost-effective.

The two main options for accessing a sample frame would be to sample directly from employer and worker surveys or to generate a sample frame using commercial samples or free-find techniques. Sampling directly from a survey is preferable in this case since it offers the advantage of triangulating survey and qualitative findings.

Qualitative samples are typically much smaller than quantitative samples and the robustness of a qualitative sample depends on the extent to which it captures the range and diversity of characteristics of the sampled population. The sample composition for both groups would need to take into account primary characteristics such as region, sector and firm size, socio-demographic characteristics and where possible relevant survey findings such as extent of knowledge or understanding and contact with LME promotional materials.

## 4.2.5  Estimated costs

Table 4:1 presents estimated timelines for the different research options on the impact of public communications and gives an indication of how expensive each option is expected to be, ranging from tens of thousands of pounds to a million pounds or more.

| Table 4:1 | Estimated costs for public communications research | |
| --- | --- | --- |
| **Approach** | **Estimated timeline from commissioning** | **Estimated Cost** |
| Randomisation to assess the impact of promotion activities on voluntary compliance<br>**Administrative data only** | 3 to 4 years | ☐ £<br>☒ ££<br>☐ £££<br>☐ ££££ |
| **With data collection** | 4 to 5 years | ☐ £<br>☐ ££<br>☒ £££<br>☐ ££££ |
| Telephone survey of employers | 9 to 12 months | ☐ £<br>☒ ££<br>☐ £££<br>☐ ££££ |
| Random probability omnibus survey of individuals | 6 to 9 months | ☒ £<br>☐ ££<br>☐ £££<br>☐ ££££ |
| Depth interviewing: Employers | 6 to 9 months | ☒ £<br>☐ ££<br>☐ £££<br>☐ ££££ |
| Depth interviewing: Individuals | 6 to 9 months | ☒ £<br>☐ ££<br>☐ £££<br>☐ ££££ |

**Randomisation to assess the impact of promotion activities on voluntary compliance**

Randomised controlled trials tend to have very high costs, often in the millions, particularly due to the time incurred in project management. The cost indication in Table 4:1 assumes however that the cost of implementing the promotion activities is shouldered by the relevant enforcement body and that the IDBR access is facilitated by Government. If a commercial dataset of businesses needs to be purchased, this will incur additional costs.

As highlighted in section 4.2.1, the approach could potentially utilise administrative data only. This is a significantly cheaper option, and would allow the trial to cover a much larger number of businesses but would potentially underestimate impact if employers self-correct without reporting this to a relevant body.

An approach using primary data collection, which is associated with higher costs than using administrative data, would likely need to be facilitated by a web or telephone survey. Such an approach would incur the costs of questionnaire development, programming, telephone interviewing and additional project management. While the additional expense may necessitate reducing the size of the trial, a smaller trial would

affect statistical power; whereby including a larger sample of businesses would increase the chances of being able to detect an impact.

**Telephone survey of employers**
A telephone survey could be commissioned and conducted relatively quickly compared with other research designs explored in this report. Again, survey costs would depend on the desired sample size. The assumption is that a sample of 2,000 businesses is required, with an estimated response rate of 20-25%. Larger samples would increase cost. Similarly, an interview length of 15 minutes should allow for a reasonable number of questions to be included. However, longer survey length would impact on both costs and response rates.

The costs for a telephone survey would include time for questionnaire development, CATI (computer-assisted telephone interview) programming, telephone interviews and interviewer briefings. The costs may be reduced if the survey were re-run with the same questionnaire. It may also be possible to incorporate the proposed qualitative research into a single project, which may make some efficiencies.

**Random-probability omnibus survey of employees**
This approach is likely to be the fastest to implement and the lowest cost of the quantitative research options outlined in this report. Omnibus panel surveys are run regularly (usually monthly) with a sample that has already been selected. This means that once the questionnaire is developed, fieldwork should be able to take place relatively promptly.

The omnibus panel would include roughly 10-15 minutes of questions and a sample of approximately 2,000 respondents. Longer surveys or larger samples would incur greater costs. This approach could also be combined with the qualitative research outlined above, which may introduce some efficiencies.

**In depth qualitative research with employers and individuals**
The indicative costs in Table 4:1 are based on samples of 60 employers and 60 workers participating in hour-long in-depth telephone interviews and assumes thank you payments will be paid to both groups to compensate for their time.

# 4.3   Direct action

Direct action incorporates both complaint-led or 'reactive' enforcement and targeted 'intelligence-led' enforcement. There are several research questions that could be explored in this area, such as:

- Does enforcement increase compliance with regulations amongst similar firms?
- Do the additional powers LME bodies have in England and Wales affect compliance?
- How does enforcement improve outcomes for workers?

## 4.3.1  Complaint-led  enforcement

Several outcomes can be considered when evaluating whether enforcement improves outcomes for workers. However, regardless of the outcome, all evaluations of complaints-led activities would likely face a common barrier in the coverage of available data. Complaint-led enforcement is recorded using data collected by the three enforcement bodies which only has coverage where complaints have been raised. Establishing a counterfactual is not possible if there is no data on a potential comparison group.

Other existing data sources, such as the Annual Survey of Hours and Earnings (ASHE) and the Labour Force Survey (LFS) are unlikely to be suitable alternatives for analysis because:

- Prevalence of workers who have been subject to a breach in labour market regulations is expected to be low

- Data collected in these surveys does not include items covering many of the aspects of interest; e.g. pay deductions and employment rights[18].

An alternative quantitative approach could utilise data collected through a respondent-driven (RDS) sample survey, such as the approach outlined in Chapter 5. This survey would have coverage of the relevant population and would already include many relevant items in the survey questionnaire. However additional items would need to be added to facilitate an evaluation of complaint-led enforcement. If items could be added be included in the questionnaire, they may want to ensure the questionnaire includes items that assess:

- Gross pay and hours

- Deductions from pay (amounts and reasons) and provision of employment rights

- Whether the individual works for an employment agency

- If the individual found their current job through an employment agency, did they pay any fees and if so, what were these fees for?

- Whether the individual had raised a complaint (e.g. not paid minimum wage) with an employer/agency. If so, was it resolved and how quickly?

- Whether the individual had raised a complaint with an enforcement body. If so, was it resolved and how quickly?

- If the individual works in a sector regulated by the GLAA, do they know if their employer holds a licence?

With these additional items, it may therefore be possible to compare the difference in outcomes for the group who have raised a complaint and those that haven't.

However, including additional items in this survey to facilitate an evaluation of complaint-led enforcement has several potential limitations. For example, there is no guarantee that there will be a large enough sample of individuals that have and haven't reported complaints to make quantitative analysis feasible as the sample is determined purely by respondents.

Furthermore, increasing survey length is known to reduce response in all surveys, though it may have further impacts for a survey with this design. A key element of respondent-driven sampling surveys is that respondents recruit other respondents. Whilst respondents who recruit others are heavily incentivised, lower response rates, particularly at the early stages of the survey when the sample is small, could be detrimental to the overall survey recruitment.

To get the required information for an item may take several questions and may need to be repeated if there are employment spells with multiple employers. Even then, there is no guarantee of quality, particularly if the respondent themselves has a poor understanding of labour market regulations and this could be challenging for interviewers. It may therefore not be possible to include all of these questions for each evaluation design.

---

[18] Such as rest breaks, paid holiday, itemised payslips including detail of deductions, sick pay and provision of terms and conditions of employment.

In summary, no design for establishing a causal impact is feasible for complaint-led enforcement, due to issues with data. It may be possible to make a quantitative assessment of some aspects of labour market regulations but doing so would be challenging. The indicative costs reported at the end of this section assume that the RDS survey is also being conducted and this therefore represents the expected additional costs of adding items to the survey and analysis of the data.

## 4.3.2 Targeted enforcement

There are several research designs that could be employed to help understand the impact of targeted enforcement and this section explores them in turn. However, many designs are not possible because of data availability issues. All options are explored in this section, but where evaluation is not currently possible because of data issues, this is clearly stated.

### Analysis of NMW and NLW non-compliance in industries targeted by the HMRC NMW team

To assess the impact of targeted enforcement on non-compliance with the National Minimum Wage (NMW) and National Living Wage (NLW), an evaluator could use difference-in-differences (DID) or an instrumental variables (IV) regression. These designs could work if the HMRC NMW team undertook targeted enforcement in a specific sector that could be identified using Standard Industry Classification (SIC) codes, which is the example explored here.

**Difference-in-differences**
DID could be used to assess the impact of targeted enforcement on the likelihood workers within those industries are not paid the minimum wage.

In an industry-level example, a list of industries[19] subject to targeted enforcement between the first period and the second period would need to be supplied by the DLME to an independent evaluator. This would be used to create a binary measure indicating if an industry was enforced against.

To compare trends in non-compliance over time, the evaluator could use ASHE, exploring the proportion of workers in the targeted and untargeted sector not paid the minimum wage over time and prior to enforcement, to satisfy the "common trends" assumption of the DID approach.

If common trends are established, the evaluator would then compare the difference in the proportion of workers who are paid less than the minimum wage between the targted and untargeted sector in period one, and similarly for period two. Taking the difference-in-differences should then establish the overall impact of enforcement. It would also be possible to conduct the same analysis assessing the impact of enforcement on the depth of minimum wage underpayment within targeted and untargeted industries.

As theorised in the logic model (Chapter 2), voluntary compliance is an expected outcome of the enforcement bodies work. Consequently, enforcement in one industry could potentially influence the outcomes in another industry. Eliminating these confounders is a key part of identifying a causal impact and consequently this is a significant limitation to the use of the DID approach in this context and an evaluator

---

[19] If the SIC is not available, it could be derived from a list of firms enforced against.

would need to be aware of the possibility of this in analysis. For this reason, it is recommended that if this analysis were undertaken, it was accompanied by analysis assesing the likelihood of spillovers, outlined in the instrumental variable approach described below. A potential approach for assessing this is described below.

**Instrumental Variables**

To use the IV approach, a suitable instrument must be available. Instruments must relate to the likelihood of targeted enforcement, but not to the outcome (minimum wage violations). Selecting a suitable instrument is difficult in the UK context. Instruments used in the examples from Chapter 3 included state-level enforcement in the U.S., the per capita number of labour market inspectors in Argentinian provinces and the development of the road network that affects inspectors "arrival cost", also in Argentina.

None of these instruments are likely to work in the UK context. Although there is a slight variation in some powers between UK countries, there is not enough to instrument in the approach taken by Galvin (2016). In the UK, the network is well developed, making the "arrival cost" approach unsuitable (Viollaz, 2018). Ronconi's (2010) instrument, the number of per capita labour market inspectors, appears most suitable of the three, but given that the UK is both smaller geographically and has better transport networks, it is probably still unsuitable in the UK context.

This section explores one instrument; a human capital measure that could be derived for the purpose of this analysis. However, the choice of instrument should be scrutinised and alternative instruments could be appropriate[20].

The instrument could be constructed identifying the level of human capital within occupations, which had previously been used as an outcome in a study of occupational downgrading following maternity leave (Connolly & Gregory, 2008). To estimate the skill level within an industry (SIC) an evaluator would derive the average level of qualifications for workers in each SIC code[21].

If a list of industries subject to targeted enforcement between the first period and the second period was to be supplied by the DLME to an independent evaluator, it could be used to construct a binary measure indicating if an industry was enforced against.

The first-stage equation in the IV analysis would use this indicator of targeted enforcement. It would also include the instrument (in this case, the human capital measure in the industry) and all the covariates included in the second stage equation.

The second stage equation would indicate if a person was paid at least their relevant minimum wage in the second period[22]. The predicted values from the first-stage equation would then be included as a covariate in the second stage equation, alongside other predictive characteristics such as: gender, age, the bite of the minimum wage in their local area[23] and whether the individual works part-time or full-time.

---

[20] Although the bite of the minimum wage had been considered as a possible instrument, this was ruled out as it is likely to be associated with the outcome as well as with the likelihood of enforcement.

[21] Connolly and Gregory (2008) create this measure within occupations rather than within industries. For this example, constructing the measure within industries is logical, but the instrument could also be conducted at an occupation level. Qualifications are assessed on an ordinal scale and the average is defined within each code, making the instrument a continuous variable.

[22] As this varies dependent on age and is also lower for apprentices. This can be identified through the data sources proposed.

[23] This could be derived from ASHE using a similar methodology to Dolton, et al. (2015).

**NatCen Social Research** | Scoping Study

The coefficient against the instrumented-variable (i.e. the predicted values from the first stage equation used as a covariate in the second stage equation) would indicate the association between minimum wage violations and working in an industry that is likely to be subject to targeted enforcement. There is also a possibility that this analysis could also explore the impact of targeted enforcement on the depth of minimum wage violations.

## Summary

Overall, this is a complex approach that could provide some indication about the impact of targeted enforcement. It utilises existing data sources, making it less costly and requiring a shorter timeline than other appproaches. However, there are a number of drawbacks to this approach, which include:

- Coverage of surveys used in analysis

- Granularity of SIC and SOC codes[24] and their relative sample sizes

- Self-reported income data (e.g. LFS) is less reliable than data sourced directly from employers (e.g. ASHE)

- It is difficult to isolate potential spillovers of targeted enforcement impacting on other employers in industries that have not been targeted for the DID.

- The suitability of instruments for the IV analysis

Undertaking these analyses in tandem is the best available approach, given the complexity of the nature of labour market enforcement. Although the DID method cannot eliminate all confounders, undertaking the IV analysis should provide an indication of what the likely spillover effects are likely to be, which could help with interpretation of the DID impact estimates.

## Application to other bodies

This analysis could be applied to work of other labour market bodies (e.g. the EAS) if suitable data was available. However, the prevalence of agency workers may mean that the sample sizes in ASHE are too small for analysis to be possible. Additionally, these sources do not contain data on some key outcomes for the EAS, such as deductions from pay and information on whether health and safety information is provided.

Although the EAS routinely collects information from agencies it enforces, such information is not available for agencies not subject to enforcement. If this information were available for agencies that were not enforced, it would potentially facilitate similar analysis. However, data collection in this context is not recommended because:

- The act of data collection is likely to influence behaviour (Hawthorne Effects[25])

- Non-response from non-compliant agencies is likely to be high

- Data from responding agencies would likely be subject to social desirability bias

Consequently, it is not recommended that data collection is undertaken to facilitate this analysis for other bodies.

---

[24] SOC is usually available to three digits and SIC to two digits in most surveys. Under special licence conditions, these are sometimes available at a greater level of granularity, but there is a trade-off between this and the number of individuals within each coded group. This is more problematic if targeted enforcement is undertaken against very specific groups of firms.
[25] This refers to a randomised controlled trial at the Hawthorne plant of Western Electric in Chicago in the early 1900s, investigating the impact of light levels on the productivity of factory workers. The study found that simply being observed increased productivity in the plant.

This approach would also be difficult for the work of the GLAA, for the following reasons:

- The volume of targeted enforcement cases (300) means that the analysis would lack the statistical power to detect an effect

- There is no existing dataset that captures the relevant GLAA outcomes

- Data collection is impractical and could lead to Hawthorne effects

## Difference-in-differences analysis exploring the impact of different powers between UK countries

There are some powers which have been granted to the three enforcement bodies that do not apply throughout the whole of the UK. For example, powers granted by the Immigration Act (2016) that apply in England and Wales do not always apply to Scotland[26]. With specific regard to LME Undertakings and Orders, the volumes are currently fairly low given that these powers are new and the processes, both within the bodies and for legal proceedings, are relatively untested.

In scenarios where powers[27] change over time, but there are some areas that continue to apply the current "business-as-usual", it would be possible to undertake analysis under difference-in-differences. This data collection would likely need to be run as a new stand-alone survey. This could include a wide range of outcomes such as deductions from pay or whether they have been given their contract in advance of accepting employment. A random-probability survey design could give an idea of prevalence of these issues, but if this figure is low, it may be desirable to use an alternative approach, such as respondent driven sampling. Using DID would also only be possible if there is a large enough sample size and variation in the outcome (i.e. a proportion of cases that are non-compliant).

DID could be feasible for analysing the impact of other powers if they were introduced in the future and didn't apply to all areas of the UK. For the approach to work, an evaluator would need outcome data (and data to be used as covariates in analysis) for at least two periods prior to the introduction of the new powers and for one period after. This presents several significant challenges:

- Data collection for some outcomes would be challenging, particularly if it requires collecting data on people whose labour market rights have been breached who are not in contact with the relevant LME enforcement bodies.

- Anticipating changes in powers is difficult as they are subject to a political process, meaning that there is a high risk that data collection could not be undertaken with enough time prior to their introduction to facilitate such analysis.

- Changes in the powers given to LME bodies may be a small change from the "business as usual". When anticipated impacts are small (even if a small impact is meaningful), more statistical power is required to detect it. It may not be possible to achieve the required level of statistical power with a sample that is achievable with data collection[28].

Starting data collection could facilitate future evaluators, but for the reasons set out above, this is anticipated to be a risky strategy. Furthermore, as data collection would

---

[26] For example, the LME Orders on application, which does not currently apply in Scotland.
[27] Throughout this section the report focuses on changes in powers, but the methodology could be applied to other scenarios. For instance, a significant change in funding or strategy, so long as there is a group that is unaffected by the change.
[28] Detecting smaller impacts requires a larger sample, which is challenging in the context of labour market enforcement if new powers were targeted at issues with low prevalence.

likely be targeting hard-to-reach groups, data collection would come at a high cost. Therefore, undertaking any data collection in anticipation of changes in the powers of enforcement bodies would have to be carefully considered.

## 4.3.3 Costs

Table 4:2 presents estimated timelines for the research designs on the impact of direct action and gives an indication, where possible, of how expensive each option is expected to be.

| Table 4:2 | Estimated costs for direct action research | |
|---|---|---|
| **Approach** | **Estimated timeline from commissioning** | **Estimated Cost** |
| Analysis of NMW and NLW non-compliance in industries targeted by the HMRC NMW team, using DID and IV | 18 months to 2 years | ☒ £<br>☐ ££<br>☐ £££<br>☐ ££££ |
| Difference-in-differences analysis exploring the impact of different powers between UK countries | Not currently possible. Would require prospective data collection. | n/a |

**Analysis of NMW and NLW non-compliance in industries targeted by the HMRC NMW team**

This design involves no data collection, instead drawing upon existing data sources. However, the ASHE data in particular is only available under secure access conditions and the analysis itself is relatively complex. This option would therefore take roughly 18 to two years to complete.

As indicated in Table 4:2, the cost of this option is expected to be in the tens of thousands. An analysis project of this length and complexity would require the research team to work relatively closely with the Office of the Director of Labour Market Enforcement to discuss the methodology and design at an early stage in the project and present findings at timely points throughout.

When data is held in secure access, it often presents an additional administrative burden, and incurs additional project management costs. Some of the variables outlined in the proposal, such as the human capital instrument, involve complex derivation and incur additional data management time. The data analysis is also complex and significant work should be undertaken to test the underlying assumptions of the approach and in conducting robustness checks.

Overall, this option could provide evidence in a medium-term horizon, at relatively low cost. Whilst it could be possible to commission the two strands of the research separately (i.e. the DID approach and the IV approach), it has been presented as one project in this report. This is because it is anticipated that much of the data management and analysis will overlap and synthesising the findings of the two strands may be more informative. The combined cost of commissioning two studies would also likely be higher than running as a single project, as the latter approach produces efficiencies.

**Difference-in-differences analysis exploring the impact of different powers between UK countries**

An overall estimate of cost is not provided for this approach. As highlighted in section 4.3.2, this approach is associated with high-risk due to the need to collect data in anticipation of change. It is likely that collecting such data on a regular basis could be a relatively expensive option. The associated data collection costs would depend on whether data is collected for employers or individuals, how much data is collected and how frequently. Due to the large scale of uncertainty about what data would be collected, it is not prudent to estimate cost.

However, whilst the analysis is dependent to some extent on the context of the intervention, it is possible to provide a rough estimate of the costs. If the data that has been collected is cleaned and easily available to the evaluator, the cost of undertaking the analysis would most likely be in the tens of thousands.

# 4.4   Prevention

The 'prevent' strand of the LME agencies' work is aimed at developing tailored guidance and protocols through engagement with industry stakeholders. This work is intended to encourage voluntary compliance including among employers that LME agencies do not have direct contact with, and to improve employer understanding of labour market regulations.

The nature of the first outcome of interest, voluntary compliance among employers that LME agencies do not have direct contact with, is difficult to assess quantitatively or qualitatively because these employers are not known to LME agencies. Low expected sample sizes, particularly for EAS engagement activities and ethical factors may also pose limitations to quantitative research in this area. For this reason, qualitative research may be best placed to explore employer understanding of labour market regulations from the perspectives of LME staff and relevant stakeholders.

## 4.4.1  Qualitative case studies

A qualitative case study design would provide an in-depth exploration of LME bodies' stakeholder engagement and co-design work. It could explore perceptions of LME staff, and relevant stakeholders of whether and how 'prevent' work affects employer understanding of labour market regulations and perceived impacts on employer behaviour.

The design of each case study would vary due to the different approaches to stakeholder engagement taken by the LME agencies, or to fit with the work being done with different sectors and the number of participating stakeholders. The case study design would nonetheless consist of:

- In-depth interviews with LME staff leading the strand of work
- In-depth interviews with participating employers and industry stakeholders
- Observations of consultation events
- A review of relevant documentation, such as draft and final guides and protocols
- Longitudinal interviews with employers and industry stakeholders

The first data collection exercise (interviews, observations and document review) would capture views, experiences and details of the process across the range of informants and detailed insights about what their work is intended to achieve. Longitudinal interviews with employers and stakeholders, conducted at an agreed timepoint, would explore perceived impacts on employer understanding and behaviour.

## 4.4.2 Estimated costs

| Table 4:3      Estimated costs for prevention research | | |
| --- | --- | --- |
| **Approach** | **Estimated timeline from commissioning** | **Estimated Cost** |
| Case study approach (12 case studies) | 12 -18 months (depending on when longitudinal interviews would be most appropriate) | ☒ £<br>☐ ££<br>☐ £££<br>☐ ££££ |

The estimated costs in Table 4:3 are based on 12 case studies in total, with interviews lasting an hour. Costs assume that data collection is conducted face-to-face and is supplemented by telephone interviews, depending on informant availability on the day of observations. Thank you payments should be considered for harder-to-engage groups such as employers.

# 4.5   Estimating non-compliance

This chapter has so far discussed a range of evaluation methods, designed to assess the impact of public communication activities, study the efficacy of complaint-led approaches to labour market enforcement and explore employers' understanding of their responsibilities.

This section addresses the issue of non-compliance with labour market regulations. Why does this arise and how extensive is it? The answers to these questions would enable the three regulatory bodies not only to fine tune their prevention strategies focusing on specific forms of non-compliance, but also provide baseline estimates which are fundamental to an understanding of the continuing impact of the work of the three regulatory bodies.

## 4.5.1 Why does non-compliance arise?

As discussed in Chapter 1, a variety of forms of labour market non-compliance exist, driven by ignorance of relevant legislation on the part of employers and/or workers; collusion between employers and workers; and exploitation of workers by employers. Non-compliance with labour market regulations therefore requires there to be a working relationship between an employer and a worker. This raises the question as to how non-compliance can be measured. Should the employer be approached for information about non-compliance, or should it be the worker, or both? Where the underlying cause is ignorance of legislation, employers may be prepared to outline their understanding of legislation, indicating where there are gaps in their knowledge. However, where the cause is collusion or exploitation, employers and possibly their workers are unlikely to reveal the nature and extent of their non-compliant behaviour.

This section outlines specific approaches which address different forms of non-compliance relevant to the work of particular labour market enforcement bodies, ranging from ignorance to exploitation. An indication of the required resources, duration and the risks associated with each of these examples is detailed below.

## 4.5.2 Example 1: Gig economy couriers – non-compliance with pay, hours and contractual conditions of employment

There are growing concerns[29] about the rising number of people employed as couriers in what is termed the 'gig economy', often on a self-employed basis and accepting work from application platforms operated by a small number of companies. Little systematic evidence exists about their working hours, pay, national insurance status, costs of maintaining equipment, etc., apart from that which is collected via the Labour Force Survey. A new approach to the collection of relevant information about this group of workers could be obtained by noting that many such workers receive work assignments via smart phones.

A focus on this group of workers may appear to be beyond the scope of the labour market enforcement bodies, given the predominance of their status as self-employed workers. However, it is increasingly being recognised that most workers in this sector have what is termed 'dependent contractor' status and will benefit from certain rights and obligations[30]. Monitoring of this group, who are placed at the margin of the sphere of responsibility of the labour market enforcement bodies, will provide evidence about the varying working conditions for those who are covered by relevant legislation as opposed to those who are currently not covered.

An internet-based survey could be designed, to be sent out via social media and targeting couriers who receive their work assignments via mobile applications. This would avoid the need to use employers as gatekeepers, collecting data directly from workers in the sector. The design of the survey would have to be given careful consideration, to ensure that it was attractive to potential respondents, short enough so as not to discourage response and ethical in its approach to data collection and data security. Significant efforts would have to be made to engage potential respondents via social media outlets (Twitter, Facebook, Instagram) and care taken to minimise false reporting.

The benefits of such a survey are that it could assist with work to ensure that, for employees, minimum wage legislation was enforced in this sector. Given the high proportion of workers in this sector who are deemed to be self-employed, this investigation could also provide evidence to indicate the extent to which employers are avoiding paying workers at or above the national minimum wage by engaging them on a self-employed basis. This approach would incur costs in the hundreds of thousands of pounds and the work involved, from design to reporting, could be completed within one year. The risks associated with this approach relate to the absence of a population sampling frame, thereby complicating the process of statistical inference from the survey findings and the possibility of falsification of responses. However, these risks are outweighed by the low costs of the study design.

## 4.5.3 Example 2: Agency workers and non-compliance with Agency Worker Regulations

The Agency Worker Regulations 2010 guarantee agency workers equal pay and conditions with employees in the same business (who do the same work) following

---

[29] See for example https://www.bikecitizens.net/food-couriers-employment-gig-economy/

[30] For example, the right to receive a written statement of their employment particulars. See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/767793/RPC4214_2_-BEIS_Modern_working_practices_-_written_statement_of_employment_particulars_for_workers_-_IA_f__-_opinion.pdf.

completion of a qualifying period of 12 weeks. Alternatively, agencies may ask an agency worker to agree to what is termed the 'Swedish Derogation'. This provides an exemption to the right to equal pay that an agency worker should receive under the Agency Worker Regulations. Instead, the worker is eligible for pay between assignments. Evidence has been presented (Metcalf, 2018) that indicates that this derogation is being used inappropriately, either through agencies offering assignments to workers that are so unattractive that they decline them or by exploiting a worker's lack of understanding of the regulations.

Some information about agency work and payments to workers can be obtained from the Labour Force Survey[31], including details such as where workers are concentrated (e.g. by sector, occupation, geographical location) and their pay, but it does not yield information about compliance with agency worker regulations.

To investigate this issue, a survey of agency workers could be undertaken. The problems such a survey would pose relate to the identification of agency workers and to those doing the same (or substantially similar) work in the organisations in which they work. The first stage would be to compile a list of agencies supplying employment services, using sources such as the Inter-departmental Business Register and/or Companies House records. A sample of these, selected on the basis of sectors identified from the LFS, could be approached with requests to approach their agency workers. Such workers would be surveyed to collect information about their assignments, their pay and the names of the employers they are currently working for or have worked for. These employers would be approached to collect information about the pay of their employees doing the same or similar work as the agency workers they are using or have used.

This is a complex data collection exercise, but one which could yield valuable information from the various stages of contact. Though Government has announced its intention to repeal the Swedish Derogation, it would be useful to design a 'before and after study' to determine the extent of compliance with the repeal order.

There are risks associated with non-cooperation from agencies that are aware of the extent of their non-compliance with agency worker regulations, but this may be mitigated by careful design of participant information literature. Approaching up to 1,000 agencies, securing cooperation from, say, 500, surveying 5,000 agency workers and 600 employers would incur very high costs, of up to a million pounds, and could take between one year and eighteen months to complete.

## 4.5.4 Example 3: Exploitation of workers covered by Gangmasters (Licensing Conditions) Rules 2009

Licensing Standard 3 of the GLAA Licensing Standards (GLAA, 2018) states that a licence holder must not:

- Subject a worker to physical or mental mistreatment

- Subject a worker to sexual violence

- Make threats to worker or others

- Abuse a worker because of any vulnerability and must pay due regard to the Equality Act (2010)

- Deceive a worker about the nature of the work, pay or living conditions

---

[31] The LFS includes this question put to agency workers: 'Are you paid on a continuous basis (between assignments)?'.

- Subject a worker to isolation, or

- Subject a worker to abusive working, living conditions or excessive overtime.

While inspections associated with licence applications and with the maintenance of these conditions undoubtedly reduce the extent of and the exploitation of workers who may be employed in violation of these conditions, there remains the possibility that some such violations may occur. Arrests and some prosecutions have been reported on the GLAA website for suspicion of acting as an unlicensed gangmaster. It is also possible that unlicensed gangmasters may be operating if they take the view that they would not gain a licence, or if they have had a licence rejected or revoked but went on to supply labour in non-regulated sectors.

This is a difficult form of non-compliance to detect and measure, given that employers must be aware of their non-compliance and would not wish to reveal that fact. Information can only be obtained from those employed by licensed gangmasters or unlicensed gangmasters operating illegally. As was discussed in section 3.2, the best approach to survey sample design in this situation is respondent driven sampling[32]. The technique requires that a small group of workers are identified who are working in the sectors covered by the gangmaster licensing conditions. Workers in this group are recruited into the sample and rewarded for their participation. They are then asked to recruit similar workers known to them through their networks. For each person recruited they are rewarded further. This process then continues with the new recruits.

There are significant risks associated with this technique, particularly to ensure that the incentives to participate do not encourage false reporting, but it provides a useful approach to the identification and estimation of the extent labour market non-compliance among workers who may be experiencing labour market exploitation. Piloting of the technique would reduce this risk, but even a pilot study, possibly focused upon a sector which is deemed to be most at risk of worker exploitation, would incur high costs, though perhaps not as high as Example 2. Achieving a sample of 2,000 workers could take up to 18 months to complete. Larger achieved samples would be proportionately costlier and longer.

## 4.5.5  Estimated costs

| Table 4:4  Estimated costs for estimating non-compliance | | |
|---|---|---|
| **Approach** | **Estimated timeline from commissioning** | **Estimated Cost** |
| Example 1: Food delivery workers | 12 months | ☐ £<br>☒ ££<br>☐ £££<br>☐ ££££ |
| Example 2: Agency workers and non-compliance with Agency Worker Regulations | 12-18 months | ☐ £<br>☐ ££<br>☐ £££<br>☒ ££££ |

---

[32] Respondent-driven sampling (RDS), combines 'snowball sampling' (getting individuals to refer those they know, these individuals in turn refer those they know and so on) with a mathematical model that weights the sample to compensate for the fact that the sample was collected in a non-random way. The technique was developed in 1997 and further elaborated in 2002, 2004, 2008 and 2009. See Appendix H for references.

NatCen Social Research | Scoping Study

| Table 4:4 | Estimated costs for estimating non-compliance | |
|---|---|---|
| Example 3: Exploitation of workers covered by Gangmasters | 18 months (2,000 respondents) | ☐ £<br>☐ ££<br>☐ £££<br>☒ ££££ |

# 5  Conclusions

This report presents the findings from a scoping exercise aimed at designing research to evaluate the work of the three labour market enforcement bodies that fall under the remit of the Director of Labour Market Enforcement (the HMRC National Minimum Wage (NMW) team; the Gangmasters and Labour Abuse Authority (GLAA) and the Employment Agency Standards (EAS) Inspectorate). This work concluded that it was necessary to develop a true picture of the size and nature of non-compliance in order to establish a baseline measure that can be used to monitor progress going forward.

The first step taken to identify an optimal evaluation framework was the design of a logic model setting out the activities and intended outcomes of the three bodies. By articulating what each body is trying to do and how, the logic model sets the foundations for designing an evaluation framework. Importantly, it shows that due to the broad scope and complexity of labour market enforcement, and the overlapping powers and responsibilities of each body, a single evaluation encompassing all aspects of enforcement is not feasible at this time.

The logic model identifies different causal chains leading to five distinct outcomes. The scoping study has found that the optimal research design and available data to assess the achievement of these outcomes varies between each of the three enforcement bodies, and that some outcomes are unsuitable for impact evaluation methods attempting to make a causal inference. Therefore, an evaluation framework based on a 'contribution analysis', which utilises several different methods, offers the most appropriate way of drawing conclusions about effectiveness and understanding the impact of the three bodies.

It is important, however, to consider the proposal for a single enforcement body (SEB) that would likely encompass the NMW team, GLAA and EAS[33]. With this in mind, it is not currently known what the labour market enforcement landscape might look like in the medium- to long-term but, depending on the changes in practice that result from such a proposal, could impact on the design of an evaluation framework. With incomplete information, it would appear, at this time, that an evaluation using a single methodological approach to evaluate all aspects of the current three enforcement bodies wouldn't be feasible under a prospective SEB. As such, a contribution analysis may still be the most appropriate approach.

Following the development of a logic model, two literature reviews were undertaken to understand the approaches applied in the UK and more widely to evaluate the impact of labour market enforcement and to explore methodologies that have been adopted to measure non-compliance. The three most common approaches to establishing a causal claim are instrumental variables, difference-in-differences and regression discontinuity design. The reviews also highlighted the usefulness of other quantitative approaches, such as telephone surveys and analysis of administrative data, which could prove valuable in evaluating outcomes for which causal estimates cannot be achieved.

The review focusing on estimating levels of non-compliance finds various advantages and limitations with the methodologies employed, particularly in relation to accessing hard-to-reach populations. This is important when considering the populations most at risk of being subject to non-compliance that are covered by the remits of the LME bodies. The review finds that, while data from statutory surveys could be used to monitor levels of non-compliance with minimum wage legislation for many workers,

---

[33] https://www.gov.uk/government/news/largest-upgrade-in-a-generation-to-workplace-rights-getting-work-right-for-british-workers-and-businesses

some non-compliance is likely to go unmeasured. It is possible that where a suitable sample frame exists (e.g. a list of registered employment agencies and/or registered agency workers) a focused survey could be conducted without incurring significant cost. However, where non-compliance is most likely to affect hard-to-reach groups, targeted research using a more novel methodology such as Respondent Driven Sampling (RDS) may be needed. Though the application of a contribution analysis would generate valuable insight, establishing a baseline of the scale and nature of non-compliance and continued consistent monitoring thereof is essential for evaluating the impact and monitoring the progress of labour market enforcement activities, whether carried out by the current LME agencies or by a single enforcement body.

Informed by the logic model exercise and reviews of literature, the report outlines potential options for future research on the work undertaken by the LME bodies. The proposed research options could feed into a contribution analysis to understand the work of the LME bodies as evidence gaps currently prevent a single impact evaluation. These cover the three streams of LME work: public communications, direct action and prevention, as well as suggestions for measuring non-compliance with labour market regulations. Where possible, approaches that could be used to infer causal impact have been proposed.

Once again, it is important to consider the implications of a Single Enforcement Body on the feasibility of the evaluation designs set out in this report. If a single body continues to undertake work in the same areas (i.e. action concerning minimum wage violations, employment agencies and sectors currently in the scope of the GLAA), it is anticipated that the proposed designs, which relate to the activities undertaken as opposed to being specific to a single body, would still be viable approaches. If a single enforcement body had a different remit or was granted additional powers not currently available to the three enforcement bodies, then these evaluation designs may need to be adapted and different designs may need to be considered.

To assess the impact of public communications on voluntary compliance, a randomised controlled trial of email, letter or texted "nudges", signposting employers to available resources, could be implemented. Whether this approach would need additional data collection depends on the quality of the outcome data that is collected by the three bodies. An alternative approach to assess public communications could be a telephone survey of employers. This could focus on where employers seek information on labour market regulations, as well as understanding of specific regulations themselves. Depth could be added to a survey by undertaking qualitative research with employers. Similarly, surveys of employees could be used to gauge their understanding of their labour market rights, and if the survey were repeated, how this changes over time. These latter approaches (surveys, complemented with qualitative research) could also be applied to public communications by different bodies and/or a single enforcement body.

Direct action includes complaint-led investigations of labour market infringements and targeted enforcement. Complaint-led activity is already monitored quantitatively by each of the three bodies but no robust impact evaluation design can be implemented to assess this. The only feasible way of adding value to the information available about complaint-led activity would be through additional data collection; likely through an RDS survey design. Targeted enforcement could be assessed more quickly and cheaply. Secondary analysis of the Annual Survey of Hours and Earnings and the Labour Force Survey could assess the impact of targeted enforcement on minimum wage compliance. To do this, the evaluator could conduct two pieces of analysis, a difference-in-differences design and an instrumental variables approach. At present, it is not possible to apply such methods to offences outside of national minimum wage regulations by whichever body might enforce it due to lack of available outcome data. In scenarios where powers which have been granted to the three enforcement bodies

do not apply throughout the whole of the UK, and some geographical areas continue to apply the "business-as-usual", it may be possible to undertake a random-probability or RDS survey, analysing the findings under difference-in-differences. The feasibility of this approach could however be affected by the introduction of a single enforcement body if it were to change the scope of LME powers, particularly should these powers extend uniformly across the whole of the UK.

Prevention activities, aimed at developing tailored guidance and protocols through engagement with industry stakeholders, are difficult to assess quantitatively, particularly as employer engagement varies in scale between the three bodies. The most viable approach may therefore be to conduct case studies with LME staff, participating employers and longitudinal interviews with key stakeholders alongside a review of the co-designed guidance and protocol documents. This kind of case study approach would be a viable method to assess prevention activities by different bodies and/or a single enforcement body.

When considering how best to measure non-compliance, the variety of forms of labour market non-compliance and their drivers offer important context. Labour market non-compliance is driven by ignorance of relevant legislation on the part of employers and/or workers; collusion between employers and workers; and exploitation of workers by employers. This has implications not only for the methodologies that can be applied but also when considering who best to approach for the required information: employers or employees. The recommended approaches below represent possible options that could generate new insights and feed into a contribution analysis by building on the current evidence base to support future evaluation. Three recommendations, relevant to the work of each body and applicable provided that the respective remit is being enforced, are:

- **Couriers in the 'gig economy' – non-compliance with pay, hours and contractual conditions of employment.** An online survey could be designed to collect information from couriers who are employed in what is termed the 'gig economy', to be sent out via social media, about their working hours, pay, national insurance status, costs of maintaining bicycles etc. Such a survey could provide evidence to indicate the extent to which employers are avoiding paying workers at or above the national minimum wage by engaging them on a self-employed basis

- **Non-compliance with Agency Worker Regulations.** The suggested approach would measure non-compliance with the right to equal pay that an agency worker should receive under the Agency Worker Regulations. It would consist of a survey of agency workers about their assignments, pay and employers, complemented by a survey of these employers about the pay of their employees doing the same or similar work as the agency workers they are using or have used

- **Exploitation of workers covered by Gangmasters.** New data collection would be required in the form of a survey of employees, employed by licensed gangmasters or unlicensed gangmasters operating illegally, using RDS as the best approach to survey sample design in this situation. A survey of this kind would identify the extent of compliance with the GLAAs licensing standards.

# Appendix A. Logic model information sources

## Interviews with key informants from within LME agencies and sponsoring departments

Two senior staff members from each of the bodies, as well as individuals from BEIS and the Home Office (the sponsoring departments) took part in in-depth interviews.

Interviews lasted between 60 and 90 minutes, and followed a standard topic guide produced in consultation with the DLME.

The interviews were recorded and information was organised into a framework for analysis. The framework contained headings relating the work of the bodies to the stages of the logic model.

## Document review

The following documents were reviewed and informed the logic model:

- Department for Business, Energy and Industrial Strategy, 2017. EAS Inspectorate Enforcement Policy Statement

- Department for Business, Energy and Industrial Strategy, 2018. EAS Inspectorate Annual Report 2017–18

- Department for Business, Energy and Industrial Strategy, 2018. NMW–NLW Enforcement and Compliance report 2017/2018

- Department for Business, Energy and Industrial Strategy and HM Revenue & Customs, 2018. NMW Service Level Agreement 2018–19 DRAFT

- Department for Business, Energy and Industrial Strategy and Home Office, 2018. Labour Market Enforcement Strategy 2018–19

- Department for Business, Energy and Industrial Strategy and Home Office, 2016. LME Undertakings and Orders code of practice

- Gangmasters and Labour Abuse Authority, 2018. GLAA Activity Briefing September 2018

- Gangmasters and Labour Abuse Authority, 2018. GLAA Strategy 2018–2021 for protecting vulnerable & exploited workers

- Gangmasters and Labour Abuse Authority and Home Office, 2018. Home Office–GLAA framework document

- HM Revenue & Customs, 2018. NMW Enforcement Strategy 2018.

# Appendix B. Detailed logic model

**Mission: Identify and support victims**

## Inputs

Staff

Funding

Legislation

Enforcement powers

Ministerial direction

IT systems

Links between LME bodies

Wider govern-mental and non-governmental links

*Public comms → Prevention → Direct action*

## Activities

**Awareness-raising**
LM regulations and LME activities promoted via press, trade press, social media, radio, webinars, faith groups etc to employers, employees and public

**Stakeholder engagement and co-production**
Consultation and discussion with employers and industry stakeholders. Co-design of tailored guidance and protocols

**Intelligence gathering**
- Worker complaints
- Public reports
- Intel from wider links (e.g. LME bodies, police, Home Office)
- Risk modelling
- Other data sources (e.g. ONS ASHE)

## Outputs

**Increased awareness**
among the public about signs of exploitation, of workers of their rights and how to seek help, of employers of their liabilities, of LME agencies' capabilities

**Agreed sectoral guidance and protocols**
specific to sectors and related industries

**Targeted enforcement activities**
- Warning letters
- Self-correction
- Enforcement notices
- Civil proceedings
- Penalty notices
- LMEUs and LMEOs
- License revocation
- Prohibition
- Prosecution
- Return visits and monitoring

## Outcomes

Improved understanding of LM regulations

Voluntary compliance (among those not investigated)

Improved quality and quantity of intelligence

Unlawful practices ceased

Employees recompensed

## Impact

Compliance with legislation

Workers protected

Level playing field for employers/ agencies

# Appendix C. Worked examples

NMW Promote activities

Mission: Identify and support victims

**Inputs**

Staff

Funding

Public comms

**Activities**

**Awareness-raising**
- Educational products aimed at employers and workers:
- Webinars
- Sectoral guides
- Fact sheets
- Leaflets
- Mass email and text campaigns
- Self-serve NMW calculator
- Guides and guidance on gov.uk
- On-line forums
- Commercial compliance toolkits

**Outputs**

**Increased awareness**
- Among workers of their rights and how to seek help
- Among employers of their liabilities
- Among employees, employers and public of LME agencies' capabilities

**Outcomes**

Improved understanding of LM regulations

Voluntary compliance (among those not investigated)

Improved quality and quantity of intelligence

**Impact**

Compliance with legislation

Workers protected

Level playing field for employers/ agencies

## GLAA Prevention activities

**Mission: Identify and support victims**

### Inputs

**Staff**

**Funding**

**Intelligence** from past experience and from links with other agencies

→ Prevention →

### Activities

**Stakeholder engagement and co-production**

• **Identify high risk sectors** to target Prevent resources at e.g. the construction industry
• **Convene** large construction industry employers and property developers to discuss voluntary compliance

Reconvene industry meetings to revise draft protocol at end of pilot period

### Outputs

**Agreed sectoral guidance and protocols**

• **Draft protocol** on ensuring labour standards are met in the construction industry.
• **Pilot draft protocol** with signed-up employers.

• **Protocol published** on ensuring labour standards are met in the construction industry.
• **Employers invited** to sign up.

### Outcomes

Improved understanding of LM regulations

Voluntary compliance (among those not investigated)

### Impact

**Compliance with legislation**

**Workers protected**

**Level playing field for employers/ agencies**

# EAS: Direct action

**Mission: Identify and support victims**

| Inputs | Activities | Outputs | Outcomes | Impact |
|---|---|---|---|---|

**Inputs**

**Staff**

**Funding**

**Legislation**

**Enforcement powers**

**Ministerial direction**

**IT systems**

**Links between LME bodies**

**Wider govern-mental and non-governmental links**

*Direct action*

**Activities**

**Intelligence gathering**
- Complaints from public, workers, unions, etc regarding conduct of employment agencies
- Proactive intelligence gathering based on risk indicators and partnership working with LME agencies, wider government and non-government links

**Outputs**

**Enforcement activities**
- Investigation of complaints
- Targeted investigations based on intelligence
- Staged enforcement activities
- Return visits and monitoring

**Outcomes**

Voluntary compliance (among those not investigated)

Unlawful practices ceased

Employees recompensed

**Impact**

**Compliance with legislation**

**Workers protected**

**Level playing field for employers/ agencies**

# Appendix D. Literature review 1 references

## Literature review on evaluation methodologies

Abadie, A., Diamond, A. & Hainmueller, J., 2015. Comparative Politics and the Synthetic Control Method. *The American Journal of Political Science,* 59(2), pp. 495-510.

Angrist, J. & Imbens, G., 1995. Two-stage least squares estimation of causal effects in models with variable treatment intensity. *American Statistical Association,* 90(430), pp. 431-442.

Bernhardt, A., Spiller, M. W. & Theodore, N., 2013. Employers gone rogue: Explaining industry variation in violations of workplace laws. *ILR Review,* 66(4), pp. 808-832.

Bhorat, H., Kanbur, R. & Stanwix, B., 2015. Partial minimum wage compliance. *IZA Journal of Labor & Development,* 4(18), pp. 1-20.

Card, D. & Kreuger, A. B., 1994. Minimum Wages and Employment: A case study in the fast-food industry in New Jersey and Pennsylvania. *The American Economic Review,* 84(4), pp. 772-793.

Connolly, S. & Gregory, M., 2008. Moving down: Women's Part-Time Work and Occupational Change in Britain 1991-2001. *The Economic Journal,* 118(526), pp. F52-F76.

Department for Business, Energy and Industrial Strategy, 2017. *Employment Agency Standard (EAS) Inspectorate: Annual Report 2016 – 2017,* London: Open Government Licence.

Department for Business, Energy and Industrial Strategy, 2018. *National Living Wage and National Minimum Wage: Government evidence on compliance and enforcement 2017/2018,* London: Open Government Licence.

Dolton, P., Bondibene, C. R. & Stops, M., 2015. *The Spatial Analysis of the Employment Effect of the Minimum Wage in a Recession: The case of the UK 1999-2010,* London: The Low Pay Commission.

Dolton, P., Bondibene, C. R. & Stops, M., 2015. *The Spatial Analysis of the Employment Effect of the Minimum Wage in a Recession: The Case of the UK 1999-2010,* London: The Low Pay Commission.

Fidrmuc, J. & Tena, J. D., 2017. UK national minimum wage and labor market outcomes of young workers. *Economics e-journal,* Volume 80.

Galvin, D. J., 2016. Deterring Wage Theft: Alt-Labor, State Politics, and the Policy Determinants of Minimum Wage Compliance. *Perspectives on Politics,* 14(2), pp. 324-350.

Gindling, T. H., Mossaad, N. & Trejos, J. D., 2015. The consequences of increased enforcement of legal minimum wages in a developing country: an evaluation of the impact of campana nacional de salarios minimos in Costa Rica. *ILR Review,* 68(3), pp. 666-707.

Hardy, T. & Howe, J., 2017. Creating ripples, making waves? Assessing the general deterrence effects of enforcement activities of the Fair Work Ombudsman. *Sydney Law Review,* 39(471), pp. 471-500.

HM Treasury, 2018. *The Green Book: Central Government Guidance on Appraisal and Evaluation,* London: s.n.

Low Pay Commission, 2017. *Non-compliance and enforcement of the National Minimum Wage,* London: Low Pay Commission.

Milkman, R., Luz Gonzalez, A. & Ikeler, P., 2012. Wage and hour violations in urban labour markets: a comparison of Los Angeles, New York and Chicago. *Industrial Relations Journal,* 43(5), pp. 378-398.

Ministerio de Economia, Infraestructura y Energia, 2019. *Encuesta Permanente de Hogares (EPH) Contínua.* [Online]
Available at: http://deie.mendoza.gov.ar/#!/documentos-metodologicos/encuesta-permanente-de-hogares-eph-continua-66
[Accessed 2019 02 26].

Parker, C. & Nielsen, V. L., 2011. Detterence and the impact of calculative thinking on business compliance with competition and consumer regulation. *The Antitrust Bulletin,* 56(2), pp. 377-426.

Ronconi, L., 2010. Enforcement and compliance with labor regulations in Argentina. *Industrial and Labour Relations Review,* 63(4), pp. 719-736.

Viollaz, M., 2018. Enforcement of labor market regulations: heterogeneous compliance and adjustment across gender. *IZA Journal of Labor Policy,* 7(2), pp. 1-28.

Vosko, L. F. et al., 2017. The compliance model of employment standards enforcement: an evidence based assessment of its efficacy in instances of wage theft. *Articles and Book Chapters,* Issue 2608.

# Appendix E. Literature review 1 search terms

## Literature review on evaluation methodologies

- National Minimum Wage (NMW)/National Living Wage (NLW)
- Labour Market Enforcement
- Labour Market Bodies/agencies
- Quasi-experimental design (QED)
- Gangmasters and Labour Abuse Authority (GLAA)/Gangmasters Licensing Authority (GLA)
- HMRC National Minimum Wage Team
- Employment Agency Standards (EAS) inspectorate / (EASI)
- Minimum wage
- Employment agencies / agency workers
- Labour (/labor) providers
- Worker/workers' rights
- Labour (/labor) standards
- Labour (/labor) abuse/exploitation
- Evaluation
- Enforcement
- Compliance
- Performance metrics
- Policy impact

# Appendix F. Literature review 2 summary

Summary of methodologies for measuring compliance/non-compliance

| Study/reference | Methodology/findings | Learning points |
|---|---|---|
| **Research using household surveys** | | |
| Ritchie et al, (2016) (UK) | Study of non-compliance with NMW among apprentices using data from Apprenticeship Pay Survey (APS) – a targeted survey of apprentices capturing self-reported income – and the Annual Survey of Hours and Earnings (ASHE), for which detailed information on income is sought from employers for a random sample of workers. The researchers compared rates of non-compliance with minimum wage rates between the two sources of data and found much higher non-compliance rates in the APS. | The research found:<br><br>• APS much more likely to capture non-compliance than ASHE, but timing may be affected by 'Friction'. Employers who are more likely to fail to comply are more likely to be missed by HMRC and ASHE;<br><br>• Triangulation between the two surveys (and qualitative research) could help better understand reasons for finding non-compliance. |
| Low Pay Commission (2017) (UK) | LPC uses ASHE to estimate non-compliance with NMW/NLW. The 2017 recent report also used LFS because the change in the date that the NLW was rolled out identified a significant change in rates of non-compliance identified in the data, which needed further exploration to see whether the increase in non-compliance was genuine. | The LPC concluded that ASHE is preferable to the LFS because the use of PAYE data rather than self-reports is likely to result in less measurement error (e.g. people rounding answers, etc., in self-report data). On the other hand, the report notes that ASHE does not capture informal economy, whereas the LFS is more likely to capture at least some of the informal economy (compared to ASHE, which uses tax records from the PAYE system to sample individuals). |
| Rani et al, (2013) (Developing countries) | Research estimating levels of non-compliance with minimum wage in 11 developing countries (Latin America, Asia and Africa) and the factors that affect this. The researchers use data from household LFS and consider | While the research allowed the researchers to examine factors related to rates of non-compliance in different countries, two limitations with the approach used are: |

| | | |
|---|---|---|
| | different institutional factors (e.g. minimum age to median wage ratio, awareness raising and enforcement activities). | • Household surveys may miss hard-to-reach communities and may underestimate informal sector;<br><br>• Many low and middle-income countries have high levels of non-compliance and so inaccuracies in estimates are less of a concern (Ritchie et al, 2016) |
| Lotti et al, (2016) (Low and Middle-income countries) | Research estimating the effect of minimum wages on levels of informal employment. Use International Income Distribution (I2D2) data compiled by the World Bank to estimate non-compliance with minimum wage legislation in 59 low and middle-income countries. Data comes from household surveys. Researchers exclude the agriculture industry from their analysis as minimum wages do not apply to agriculture in many countries. | As with the above study, while the cross-country analysis used in the study allowed examination of the factors related to non-compliance, the measurement error associated with the use of household survey data may be more of an issue in UK data where levels of non-compliance are much lower than in developing countries.<br><br>In addition, it would not be appropriate to exclude data from agriculture in the analysis of UK data as the minimum wage still applies to agricultural workers in the UK, unlike in many developing countries where subsistence farming is much more common. |
| Marshall (2007) (Latin America) | Research examining the factors associated with non-compliance of labour protective regulations in 15 Latin American countries using multivariate analysis of secondary data from household surveys. Competitiveness of internal and/or international markets was not found to predict non-compliance, but enforcement levels, cultural norms, and historical enforcement levels do predict non-compliance | Cross-country analyses use data from different countries to estimate effect of different measures (e.g. enforcement activities and contextual information) on non-compliance. While this can help understand what drives non-compliance and/or what the effects of stronger or weaker regulation is on levels of non-compliance, this approach to analysis is less useful in the case of the UK where, in most cases, the same regulations apply nationally. |
| Maloney and Nuñez Mendez (2003) (Latin America and the Caribbean) | Research using data from Inter-American Development Bank (IADB) to estimate non-compliance with minimum wages in Latin America and the Caribbean. Data is based on household surveys from nine countries. | Two limitations identified with the methodology were:<br><br>• Household surveys may miss hard-to-reach communities and may underestimate informal sector |

| | | Difficulties estimating minimum wages in absence of data on hours (may affect some industries/occupations or employment modes more than others) |
|---|---|---|
| Kristensen and Cunningham (2006) (Latin America and the Caribbean) | Research looking at the effect of minimum wages on wage distributions and informal working in 19 Latin American and Caribbean countries. Used data from Living Standard Measurement Survey (LSMS), LFS or other household surveys. | As with the above cross-country studies, focusing on non-compliance in developing countries, some of the drawbacks to this methodology in relation to the UK context are:<br><br>• Household surveys may miss some hard-to-reach communities;<br><br>• Measurement error associated with self-report data in household surveys may be more of an issue in the UK where rates of non-compliance are low;<br><br>• The methodology relies on differences in regulation between countries in order examine the effects on wages and informal working. This would not be suitable in cases where the same regulations apply to all cases/regions of study. |

| Nelms et al (2011) (Australia) | Research attempting to more accurately estimate levels of non-compliance with federal minimum wage regulations. The research used data from four sources in order to try to triangulate findings: Survey of Employment Arrangements Retirement and Superannuation (SEARS); Survey of Income and Housing (SIH); Household Income and Labour Dynamics in Australia (HILDA); and the ABS Employee Earnings and Hours (EEH) survey. Levels of non-compliance found in the different surveys ranged from 7 to 9%. | The researchers identified a number of measurement issues, many of which were related to estimating hourly earnings from weekly data. <br><br> The researchers identified five reasons for finding wages below the minimum wage for individuals: <br><br> 1. exclusions (e.g. training wage); <br> 2. long hours with no paid overtime in reference week; <br> 3. does not account for non-wage benefits (e.g. pension contributions); <br> 4. genuine non-compliance (e.g. pay below the relevant rate or illegal deductions); <br> 5. Measurement error (these are more of a problem in cases where estimated non-compliance is low – 7-9%). |
|---|---|---|
| **Targeted surveys** | | |
| Weil (2005) (US) | Survey of employers in apparel industry in Los Angeles. Randomly selected (inc. boost of known violators), sampled from Californian Manufacturers Register, weighted by assumptions about estimated non-compliance, 'inspection-based survey' (payroll review). The research found that certain regulatory levers could be used as an effective deterrent to non-compliance: the move to LEAN methods among organisations in the industry meant that the ability to embargo goods is an effective deterrent. | The methodology used allowed the researchers to identify levels of non-compliance among employers in the apparel manufacturing industry, including non-compliance with a range of regulations. However, limitations with an employer survey of this sort are: <br><br> • The survey made use of a sample frame for a known population (i.e. organisations had to be registered). This may mean that contractors/suppliers may have been missed and would not be possible for industries where no compulsory register exists; |

| | | • The researchers had to make assumptions about expected levels of non-compliance in order to weight the data. |
|---|---|---|
| Bernhardt (2009) (US) | Survey of 4,387 workers from low-wage industries in Los Angeles, Chicago and New York. Respondent-driven sampling (RDS): 'chain-referral sampling' but weights respondents based on size of their network (and observed biases). Used 'dual-incentive structure' whereby respondents were paid for taking part and for making referrals using a coupon system. While this approach leads to a biased sample this bias can be adjusted for by taking account of the relative size of respondents' networks, producing relatively unbiased estimates under certain conditions. Post-stratification adjustments were also made based on observed biases in recruitment patterns. The research found that rights violations were widespread including: pay below the minimum wage, unpaid overtime, off the clock working, no or interrupted meal breaks and wage theft. | The methodology used in the study allowed for the estimation of levels of non-compliance that workers from a hard-to-reach community were subjected to. The focused nature of the survey also allowed the research to identify a wide range of non-compliance with labour regulations aside from non-compliance with minimum wages. Drawbacks to the methodology used, however, include: <br><br> • The survey methodology used was labour intensive and costly; <br><br> • It was necessary for the researchers to know some population characteristics to enable weighting of the data (this could lead to bias if reliable population data is not available); <br><br> • The researchers emphasised the importance of communicating eligibility criteria to respondents. Difficulties in the ability to communicate eligibility criteria to respondents could lead to bias in the survey sample. |
| Friberg (2012) (Norway) | The author employed RDS to survey a population of Polish migrants in Oslo, Norway. The research aimed to study 'the complex relationship between large-scale labour migration and structural changes in the labour market and the ways in which these changes shape the opportunity structures that face migrant workers' (p5). Reasons for using RDS were that for Polish migrants: | Benefits of the use of RDS were seen as: <br><br> • Utility in accessing an otherwise hard-to-reach population; <br><br> • Allows for unbiased estimates and generalisation (compared to convenience or normal snowball sampling); |

| | | |
|---|---|---|
| | 1. no complete sampling frame exists,<br><br>2. they exhibit high levels of mobility,<br><br>3. many live in unconventional housing, many work long hours,<br><br>4. possible concerns about legal status might make some reticent to take part in ordinary surveys.<br><br>Interviews were carried out face-to-face in Polish (n=501). | • Design effects can be estimated;<br><br>• Some methods of validation are possible such as 'computational validation' (see Heckathorn) and 'empirical validation' (e.g. comparing certain characteristics against previous rounds of the survey or compared to external population data).<br><br>One of the issues that was noted in the research was that incentives that are too high can result in 'false respondents', although in the case of Friberg's research this was unlikely as language was a marker that respondents were from the population of interest. |
| **Administrative data and targeted enforcement action** | | |
| FWO (2018a) (Australia) | In response to intelligence from a variety of sources, the FWO launched an in-depth inquiry into workplace arrangements along the 'Harvest Trail' in Australia. The aims were: to understand employment and market conditions; to identify drivers of non-compliance; to educate growers, labour hire contractors and workers about workplace rights and obligations; to address non-compliance by taking compliance and enforcement action where necessary. The inquiry involved: stakeholder engagement; targeted awareness raising activity and educational communications; consumer research; and compliance and enforcement activities. The latter involved FWO inspectors visiting farms across Australia interviewing workplace participants, investigating employers and taking enforcement action as appropriate. In the first two years of the inquiry FWO inspectors devoted considerable time to informing employers of Fair Work Act rules and coverage and promoting tools to aid compliance (inc. 1,295 educational and compliance | The inquiry enabled the FWO to conclude that non-compliance was widespread, to understand how non-compliance comes about in that specific context and what the drivers of non-compliance are.<br><br>The inquiry, however, does not allow the FWO to estimate levels of non-compliance in the sector more widely as investigations were targeted at employers where concerns about non-compliance were raised (i.e. the sample is likely to be biased towards non-compliant firms). |

| | | |
|---|---|---|
| | interactions and 836 completed investigations). Visits were often in response to intelligence (e.g. requests and/or tip offs). Where substantial non-compliance was found, the region was revisited the following year. In cases where minor non-compliance due to error was identified employers were given the chance to rectify it voluntarily. More serious non-compliance led to formal cautions and infringement notices, with further breaches leading to escalation. Overall, 638 employers were investigated (inc. 444 growers and 194 labour hire contractors). More than half (56%) of investigations found a failure to comply with employment regulations (e.g. underpayment or a failure to keep records or issue payslips). The inquiry also helped identify challenges to enforcement such as poor record keeping, invalid or absent paperwork, reluctance from workers to share information and transient workers and labour hire contractors. | |
| FWO (2018b) (Australia) | Responding to concerns that security workers in the local government sector were being underpaid, the FWO launched an inquiry focusing on the procurement arrangements in 23 local governments across all seven states in Australia. As with the Harvest Trail inquiry the work involved education and awareness campaigns, engagement with employers and stakeholders such as trades unions and industry and employer bodies. The FWO carried out site visits and audits with 49 businesses, interviews with employers, employees, and requested and reviewed wage records. The inquiry found that every state (except the Northern Territories who employed security staff directly) had at least one incidence of non-compliance. Non-compliance was more common where employment relationships were more distant: 42% of principal contractors failed to comply and 63% of | As with the Harvest Inquiry (above), because organisations covered were investigated in response to concerns and were not sampled randomly, it is hard to know whether they are representative of local councils more widely and therefore it is difficult to estimate overall levels of non-compliance more widely. However, given that there are a finite number of such organisations and that the councils investigated were subject to advanced warning about the inquiry and were involved in educational and awareness raising activity (and many even had in-house legal advice in the procurement process) the fact that so many had non-compliance in their supply chain is perhaps surprising. |

| | | |
|---|---|---|
| | subcontractors failed to comply with wage regulations. Overall, 61% of councils had non-compliance in their supply chain and 12% of employee records checked had some underpayment. Reasons for underpayment included: below minimum wage hourly rate, non-payment of penalty rates, overtime and issues around part-time hours. | |
| PwC (2012) (Australia) | The FWO commissioned Price Waterhouse Cooper (PwC) to investigate 'Phoenix Activity' (where organisations deliberately and systematically go into liquidation in order to avoid tax and other liabilities – e.g. employee entitlements). The aims of the research were:<br><br>1. to define Phoenix Activity,<br><br>2. quantify Phoenix Activity, and<br><br>3. to identify possible actions to address Phoenix Activity.<br><br>The research did not involve any primary research but estimated the impact of Phoenix Activity using a 'risk-based' model based on previous research and extrapolating from secondary data from the General Employee Entitlements and Redundancy Scheme (GEERS). | Estimates of the extent of 'Phoenix Activity' are likely to be highly speculative. However, in the absence of any other data on the practice, some estimation of the likely size and impact of the practice was felt to be warranted. The researchers carried out a sensitivity analysis using different methods of estimation in order to provide some indication of likely lower and upper-bound estimates. While this gives some indication of the uncertainty around their estimates it does not overcome the issues around trying to extrapolate estimates from one source of data to a wider population. |
| Howe, Hardy and Cooney (2014) (Australia) | The FWO commissioned researchers from the University of Melbourne to carry out a review of their activities from 2006 to 2012. The researchers used quantitative and qualitative data from three sources:<br><br>1. FWO data on investigations, detection, education/media campaigns and use of sanctions;<br><br>2. qualitative interviews and participant observation with FWO staff; and | Data from campaigns allows some level of comparison of compliance rates in different industries. The high number of audits/site visits in some industries (ranging from 211 to 1,866 over 1-2 years), combined with the fact that some of these are randomly selected, opens up the possibility that these could be used to estimate/model levels of non-compliance in particular industries. Howe et al note that more research is needed to estimate compliance rates in |

| | | |
|---|---|---|
| | 3. analysis of FWO policy documents, published decisions and enforceable undertakings.

The FWO employs two methods of detection: 'complaints' (e.g. from employees) and 'campaigns' (inc. awareness, audits and site visits – inquiries are one type of campaign). Campaigns are carried out in specific industries. Some audits and site visits are targeted and in some cases they are randomly selected. The researchers report non-compliance rates in different industries, although they note that they may not always be comparable. | more depth and to model for different industries or occupations. |
| Ji and Weil (2015) (US) | Looked at levels of compliance with federal minimum wage and overtime standards in franchised and company-owned establishments in the fast food industry in the US. They found that levels of non-compliance were higher in franchised establishments. The research used administrative data from the Wage and Hour Investigation Support and Reporting Database (WHISARD), which is an administrative database that records all workplace investigations made by the US Wage and Hour Division (WHD) or the US Department of Labor. They use the data to estimate the importance of different factors (inc. ownership status) in predicting non-compliance with minimum wage and overtime standards. | Using administrative data allowed the researchers to examine which establishment characteristics were related to non-compliance. However, because the data only covers establishments that were investigated (and not those that were not investigated), it does not allow estimation of overall levels of non-compliance in the population (i.e. all fast-food establishments).

On the other hand, the research makes use of data that is collected anyway in the process of the WHD enforcement work and so there is no additional cost for collecting primary data. |
| Weil (2012) (US) | Reports on research using the WHISARD database looking at the effect of different ownership models on compliance/non-compliance with wage regulations in the fast food industry (see above) and the hotel and motel industry. Non-compliance was measured as back wages (number of employees owed and total owed). Non-compliance was found to be highest in franchise and third-party management and subcontracting type organisations. | As noted above, while the administrative data used enabled the researchers to explore the characteristics that were related to non-compliance among organisations who were investigated, as the data used only covers those who were subject to investigation it does not allow for estimation of non-compliance more widely. |

| Kleiner and Weil (2010) (US) | Use National Labor Relations Board (NLRB) administrative data on adjudications to investigate the effectiveness of penalties as a deterrent. They performed a descriptive analysis of data on numbers of citations, success rates, duration before decisions, value of awards, etc, and changes in these over time, and found that the implied penalties (i.e. re-imbursement of lost wages) were only a modest deterrent in incentivising compliance. | While the data allows analysis of changes in the effectiveness of some measures over time, National Labor Relations Act violations are only investigated if a complaint has been raised by an employer or employee and is therefore reliant on them coming forward. Therefore, data such as this cannot be used to estimate wider levels of non-compliance. |

# Appendix G. Literature review 2 references

## Literature review on measuring compliance/non-compliance

Belman, D. & Wolfson, P. (2014). What does the Minimum Wage Do? Kalamazoo: Upjohn Institute.

Fair Work Ombudsman [FWO] (2018a). The Harvest Trail Inquiry: A report on workplace arrangements along the harvest trail. Canberra: FWO.

Fair Work Ombudsman [FWO] (2018b). An Inquiry into the procurement of security services by local governments. Canberra: FWO.

Friberg, J. H. & Horst, C. (2014). RDS and the Structure of Migrant Populations. In G. Tyldum & L. G. Johnston (eds.), Applying Respondent Driven Sampling to Migrant Populations: Lessons from the Field. Basingstoke: Palgrave Macmillan. doi: 10.1057/9781137363619.0009

Friberg, J. H. (2012). The Polish worker in Norway: Emerging patterns of migration, employment and incorporation after EU's eastern enlargement. Unpublished PhD thesis submitted to the Department of Sociology and Human Geography, Faculty of Social Sciences, University of Oslo.

Ji, M. & Weil, D. (2015). The Impact of Franchising on Labor Standards Compliance. ILR Review, 68(5), 977-1006.

Kleiner, M. M. & Weir, D. (2010). Evaluating the Effectiveness of National Labor Relations Act Remedies: Analysis and comparisons with other workplace penalty policies. NBER Working Paper 16626. Cambridge, MA: NBER. http://www.nber.org/papers/w16626

Kristensen, N. & Cunningham, W. (2006). Do Minimum Wages in Latin America and the Caribbean Matter? Evidence from 19 Countries. Policy Research Working Paper No. 3870. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/8337

Lotti, G., Messina, J. & Nunziata, L. (2016). Minimum Wages and Informal Employment in Developing Countries. http://jobsanddevelopmentconference.org/wp-content/uploads/2016/10/LOTTI_Minimum-Wages-and-Informal-Employment-in-Developing-Countries.pdf

Low Pay Commission [LPC] (2017). Non-compliance and enforcement of the National Minimum Wage (September 2017).

Maloney, W. F. & Nuñez Mendez, J. (2003). Measuring the Impact of Minimum Wages: Evidence from Latin America. In J. Heckman & C. Pagés, Law and Employment: Lessons from Latin America and the Caribbean. NBER Working Paper 10129. https://www.nber.org/papers/w10129.pdf

Marshall, A. (2007). Explaining non-compliance with labour legislation in Latin America: A cross-country analysis. ILO Discussion Paper. Geneva: ILO.

Nelms, L., Nicholson, P. & Wheatley, T. (2011). Employees earning below the Federal Minimum Wage: review of data, characteristics and potential explanatory factors. Fair Work Australia.

PriceWaterhouseCooper [PwC] (2012). Phoenix Activity: Sizing the problem and matching solutions. Canberra: FWO.

Ritchie, F., Veliziotis, M., Drew, H. & Whittard, D. (2016). Measuring compliance with minimum wages.

Stewart, A. & Owens, R. (2013). Experience or Exploitation? The Nature, Prevalence and Regulation of Unpaid Work Experience, Internships and Trial Periods in Australia. Report for the Fair Work Ombudsman.

Weil, D (2005). Public Enforcement/Private Monitoring: Evaluating a new approach to regulating the national minimum wage. ILR Review, 58(2), 238-257.

Weil, D. (2012). Examining the Underpinnings of Labor Standards Compliance in Low Wage Industries (Final report). Report to the Russell Sage Foundation.

# Appendix H. Further reading: respondent driven sampling

The following references supply links to articles which describe respondent driven sampling (RDS):

Heckathorn, Douglas D. 1997. "Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations." Social Problems.

Heckathorn, Douglas D. 2002. "Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations." Social Problems.

Salganik, Matthew J. and Douglas D. Heckathorn. 2004. "Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling." Sociological Methodology.

Heckathorn, Douglas D. 2007. "Extensions of Respondent-Driven Sampling: Analyzing Continuous Variables and Controlling for Differential Recruitment." Sociological Methodology.

Wejnert, Cyprian and Douglas D. Heckathorn. 2008. "Web-Based Network Sampling: Efficiency and Efficacy of Respondent-Driven Sampling for Online Research." Sociological Methods and Research.

Volz, Erik and Douglas D. Heckathorn. 2008. "Probability Based Estimation Theory for Respondent-Driven Sampling." Journal of Official Statistics.

Wejnert, Cyprian. 2009. "An Empirical Test of Respondent-Driven Sampling: Point Estimates, Variance, Degree Measures, and Out-of-Equilibrium Data." Sociological Methodology.

Ritchie, F., Veliziotis, M., Drew, H., and Whittard, D. (2018). "Measuring compliance with minimum wages", *Low Pay Commission*. Available at http://eprints.uwe.ac.uk/35864/3/measuring%20non-compliance%20v17a%20JESM%20format.pdf

For a critical appraisal of bias and variance of RDS provided estimates, see:

Sharad Goel and Matthew J.Salganik (2010). Proceedings of the National Academy of Sciences