EVALUATION REPORT

Step Together

Pilot Report

Tom Kirchmaier, assisted by Shubhangi Agrawal, Julia Friedberg, Yasaman Saeidi and Ricky Wang

January 2025

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Home Office

YOUTH ENDOWMENT FUND

About the Youth Endowment Fund

The Youth Endowment Fund (YEF) is a charity with a mission that matters. We exist to prevent children and young people from becoming involved in violence. We do this by finding out what works and building a movement to put this knowledge into practice.

Children and young people at risk of becoming involved in violence deserve services that give them the best chance of a positive future. To make sure that happens, we'll fund promising projects and then use the very best evaluation to find out what works. Just as we benefit from robust trials in medicine, young people deserve support grounded in the evidence. We'll build that knowledge through our various grant rounds and funding activity. And just as important is understanding children and young people's lives. Through our Youth Advisory Board and national network of peer researchers, we'll ensure that young people influence our work and that we understand and are addressing their needs. But none of this will make a difference if all we do is produce reports that stay on a shelf.

Together we need to look at the evidence and agree what works, then build a movement to make sure that young people get the very best support possible. Our strategy sets out how we'll do it. At its heart it says that we will fund good work, find what works and work for change. You can read it here.

For more information about the YEF or this report please contact:

Youth Endowment Fund
C/O Impetus
10 Queen Street Place London
EC4R 1AG

www.youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

Registered Charity Number: 1185413

Contents

About the evaluator

The team that will conduct the impact evaluation is part of the Centre for Economic Performance's (CEP's) policing and crime research group at the London School of Economics and Political Science (LSE). The CEP policing and crime research group has an established track record of studying the full range of topics relating to the economics of crime. It has established itself as one of the leading groups working to gain an empirical understanding of the causes and consequences of crime through the study of administrative datasets from UK police forces. It also has extensive experience dealing with administrative data from West Midlands Police. The team has a successful record for bringing academic insights to the important problems confronting police forces in the UK.

Prof. Kirchmaier, the principal investigator for the project, has led various successful projects for the Metropolitan Police Service, the Home Office and His Majesty's Inspectorate of Constabulary and Fire & Rescue Services. As principal investigator he will be the YEF point of contact. He will coordinate the team, lead the research activities and have overall responsibility for the quality of the project outputs, the management of risks and the implementation of risk mitigation measures.

The other evaluators are:

- **Ms Shubhangi Agrawal – project coordinator (LSE). Shubhangi will be the contact point** for the impact evaluation and will coordinate the empirical side of the project.
- **Ms Julia Friedberg – policy officer (LSE).** Julia will be the contact point for the impact evaluation and will coordinate the empirical side of the project.
- **Ms Yasaman Saeidi – senior researcher (LSE).** Yasaman will assist in conducting the data analysis.
- **Mr Ricky Wang – junior researcher (LSE).** Ricky will assist in conducting the data analysis.
- **Mr Matteo Sandi from the LSE** may also provide ad hoc as a researcher, as needed. We would also like to disclose that he is an expert with the YEF.

For further information about this evaluation, please contact Prof. Tom Kirchmaier at t.kirchmaier@lse.ac.uk.

Executive summary

The project

The Step Together project aims to reduce crime and violence by placing trained adult 'chaperones' on routes used by children walking to and from school in areas identified as having a heightened risk of violence or antisocial behaviour (ASB). Inspired by Chicago's Safe Passage programme, in Step Together, the West Midlands Violence Reduction Partnership (WMVRP) commissioned 10 local voluntary and community sector organisations to provide staff members as chaperones and identified 18 routes associated with 24 schools across Birmingham, Coventry, Walsall, Wolverhampton, Solihull, Sandwell and Dudley. The WMVRP also liaised with schools to secure their buy in, conducted walk-throughs of routes to identify areas at potential risk and devised a training programme for chaperones. Chaperones were then present on selected routes; there were, on average, six on each route for at least two hours daily (at the start and end of each school day). Routes were chaperoned for between 20 and 43 weeks during the 2021–2022 academic year. In addition to being present on the routes, chaperones also interacted with pupils (generally 11-18 years old) and community members such as local shop staff, introducing themselves, enquiring about local issues and asking the young people how they were. There were approximately 18,000 pupils in the schools involved.

The Youth Endowment Fund (YEF) and the Home Office funded the delivery of Step Together from 2021–2022, in addition to funding an implementation and process evaluation and this pilot study. The previously published implementation and process evaluation found that Step Together was delivered largely as intended. The project retained the core features of the US model, and training was well received. However, adaptations were made to suit a new context, and there were some challenges in delivery (such as the selection and mobilisation of schools, routes and providers, which proved complex and time-consuming).

This pilot study aimed to ascertain whether a more robust quantitative evaluation of the programme could be undertaken. Specifically, it first explored whether the evaluator could collect the necessary crime and education data and whether they could construct a valid counterfactual to evaluate Step Together. The evaluation aimed to pilot a quasi-experimental design study in which the level of ASB and crime on Step Together routes was compared to those in adjacent areas. In addition, it aimed to compare attainment and attendance in Step Together schools with that in 171 other schools in the West Midlands that did not receive the intervention. The evaluation also surveyed participating children on their perception of the programme.

Key conclusions
There were significant delays in receiving access to the crime data required to assess the impact on crime. For future evaluations, West Midlands Police recommends that a clear specification of the data required, including specific fields, definitions, time frames and any legal agreements needed, is provided.
The evaluator aimed to use neighbouring areas as a counterfactual for measuring the impact of Step Together on crime and ASB. However, there are limitations associated with this approach, including the requirement that Step Together routes and neighbouring areas are comparable (which the pilot was not able to establish). Creating valid counterfactuals for measuring education outcomes also proved challenging.
The evaluator was unable to draw conclusions as to whether the programme showed early evidence of promise. The evaluation was never intended to evaluate the programme's impact. However, limitations in design (including a lack of evidence for the parallel trends assumption and the potential effects of the pandemic) severely limited the evaluator's ability to comment on Step Together's potential interaction with crime/ASB, student attendance and attainment.
Response rates to the pupil survey were extremely low (below 10%). As a result, the evaluators decided to discontinue the survey, and this pilot study cannot comment on pupil perceptions of Step Together.
Given the limitations of this evaluation, the YEF is not currently planning any further evaluation of Step Together.

Interpretation

There were significant delays in receiving access to the crime data required to assess the project's impact on crime. For future evaluations, West Midlands Police recommends that a clear specification of the data required, including specific fields, definitions, time frames and any legal agreements needed, is provided. The data access procedures required to analyse attendance and absence data also took a substantial amount of time; this time will need to be accounted for in future evaluations.

The evaluator aimed to use neighbouring areas as a counterfactual for measuring the impact of Step Together on crime and ASB. However, there are limitations associated with this approach, including the lack of evidence for parallel trends in ASB in the treatment and control areas. COVID-19 also caused challenges. It is likely that crime and ASB patterns were significantly impacted by the pandemic, and while the evaluator can exclude data from 2019–2021, we cannot know whether the pandemic structurally changed crime and ASB patterns. This means that even if the evaluator could determine that neighbouring areas were a plausible counterfactual pre-pandemic, we cannot know whether this remains the case after COVID-19. Identifying and creating valid counterfactuals for measuring educational outcomes also proved challenging. Future studies should consider incorporating randomisation to ensure appropriate counterfactuals. If randomisation is not feasible, it is recommended that future evaluators use a more extensive and broader dataset that contains information on observable characteristics. These characteristics can be used as controls in a quasi-experimental design to reduce confounding.

The evaluator was unable to draw conclusions as to whether the programme showed early evidence of promise. The evaluation was never intended to evaluate impact. However, limitations in design (including a lack of evidence for the parallel trends assumption and the potential effects of the pandemic) severely limited the evaluator's ability to comment on Step Together's potential interaction with crime and ASB, student attendance and attainment. However, the evaluation does provide useful advice for future quantitative evaluations of Step Together (and/or similar school-chaperone programmes).

Response rates to the pupil survey were extremely low (below 10 per cent). As a result, the evaluators decided to discontinue the survey, and this pilot study cannot comment on pupil perceptions of Step Together, including whether or not they showed increased feelings of safety. While the evaluators were not able to identify reasons for the low response rate, they advised that, in future studies, using printed surveys and having someone present in the schools to administer them may lead to higher response rates. The evaluators also attempted to design a GPS application to track the route of chaperones. However, this proved difficult to implement (due to older phones being unable to support the app and chaperones not turning the app on). This made it very difficult to establish whether the prescribed routes were used; future evaluations should explore this.

The evaluation suggests that it cost the WMVRP £1.3m to run the Step Together programme during one year. This estimate means that it cost approximately £65,000 for each route and £73 for each child (assuming the programme had an indirect impact on all 18,000 children attending the schools).

The YEF is not currently planning further evaluation of Step Together. However, the previously published implementation and process evaluation provides useful steers on how to deliver school-route chaperone programmes. Given the evidence from US-based evaluations of Safe Passage, the feasibility of Step Together, and the fact that we know that children are most at risk of involvement in violence immediately before and after school,[1] school-route chaperone programmes remain a good bet for keeping children safe. However, further robust evaluation in a UK context (likely in the form of a cluster randomised controlled trial) would be required to strengthen this evidence.

---

[1] Vulliamy, P. et al. (2018). Temporal and geographic patterns of stab injuries in young people: A retrospective cohort study from a UK major trauma centre. *BMJ Open 8*, e023114. https://bmjopen.bmj.com/content/8/10/e023114

Introduction

Background

There were 46,950 knife-enabled offences in England and Wales in 2021.[2] The UK Government's commitment to decreasing serious violence (including knife crime) led to the launch of the Serious Violence Strategy in 2018.[3] While law enforcement remained important, the government's focus with this strategy was to identify the root causes of the problem and find avenues to support young people and enable them to lead productive lives away from violence. This strategy continues to operate in 2025.

The Step Together pilot was based on evidence from a similar scheme in Chicago, Illinois, called Safe Passage. Safe Passage is run by the Chicago Public Schools with support from the Chicago Police Department. It was developed as a grassroots movement, led by the community, for the community. The programme hires, trains and places civilian guards along routes to and from schools during arrival and dismissal times. The intervention began with 35 schools in the 2009–2010 academic year and expanded to cover about 20 per cent of Chicago public schools by the 2015–2016 academic year. Schools were not randomly chosen; rather, schools in high crime, low-income neighbourhoods were chosen to be part of the programme, and guards were placed along the routes with the highest footfall. Evidence suggests that Safe Passage has been associated with an overall crime reduction of about 17 per cent and with a significant decline in violent and property crimes along the Safe Passage routes of guarded schools and in their closest neighbourhoods (Gonzalez & Komisarow, 2020). There is also evidence to suggest that Safe Passage is associated with improvements in school attendance and that Safe Passage schools have increased their attendance rates by 2.5 per cent on average when compared to other Chicago Public Schools (McMillen, Sarmiento-Barbieri & Singh, 2019).

Step Together provides a positive, trusted adult presence for students as they travel to and from school. The Step Together pilot represents a good opportunity to assess the feasibility and potential impact of a well-evidenced international model and its ability to reduce violent crime and antisocial behaviour (ASB) and improve student safety in the UK. This is the first time the project will have been piloted in the UK. To fully understand the barriers to implementation, it is important that the evaluation considers the UK context and includes the process evaluation conducted.[4]

Intervention

The market research firm Ipsos, in collaboration with the Home Office, Youth Endowment Fund (YEF) and West Midlands Violence Reduction Partnership (WMVRP) developed a theory of change/logic model to set out how the pilot for the Step Together programme was to work and its intended outcomes (Figure A.1 in the appendix). The purpose of the Step Together model was to provide a positive and trusted adult presence for students as they travel to and from school. It was piloted in the West Midlands in the academic year 2021–2022. As per the logic model/theory of change, Step Together was launched based on a series of hypotheses. The activities from this intervention were expected to lead to: a) a reduction in crime and ASB in the

[2] Office for National Statistics (2022, April 28). Knife or sharp instrument offences. *Crime in England and Wales: Year ending December 2021*. https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingdecember2021

[3] Home Office (2018, April). *Serious violence strategy*. Gov.UK. https://www.gov.uk/government/publications/serious-violence-strategy

[4] Bierman, R. et al. (2023, March). Step Together pilot: Implementation and process evaluation. Youth Endowment Fund. https://youthendowmentfund.org.uk/funding/evaluations/step-together/

identified routes, b) a potential reduction in crime and ASB in adjoining routes, c) decreased school absences and an increase in student attainment in participating schools and d) increased feelings of safety for students.[5]

The project was delivered across 18 routes in the West Midlands,[6] with an average of six chaperones per route monitoring high footfall routes for at least two hours per day at peak times, before and after school.[7] It was managed by the WMVRP, with delivery planned through local voluntary and community sector partners. As with the Chicago model, the chaperones would not play an enforcement role but were instead trained to deescalate conflict and manage safeguarding concerns. The zones within each local authority in which routes were placed were identified using the best available data: violence data from the mostly pre-pandemic academic year (2019–2020) and local knowledge from police officers, schools and WMVRP teams. The WMVRP strategic needs assessment and schools risk matrix was also used. More detailed and more current data then became available and influenced the proposed routes. The routes varied in length and included streets, roads and parks. The WMVRP team carried out walkabouts with community members and partners to identify the presence of alternative capable guardians (for example, business improvement district security staff in city centres and staff from other detached youth programmes) and the areas of higher risk. The routes were then agreed between the WMVRP, providers, police and schools based on a review of this material and observations of the students' natural routes.

Tactical information, such as the locations for chaperones on these routes, emerged from the co-design process. These decisions have required both input from schools and providers' 'soft intelligence' relating to areas in which the young people may congregate and where they feel safer. Local non-governmental organisations and community interest company youth service providers were commissioned to provide detached youth workers as chaperones on each of the routes.

Research questions

Given that this is a pilot study, we would like to understand the following, as outlined in the Step Together evaluation study plan:

1) Can we collect the necessary crime and education data within the initial time frame set out for the evaluation?

2) How can a plausible counterfactual be constructed?

3) Is there enough evidence of promise to warrant a full impact study? Ideally, an impact evaluation would address the following questions (and we will aim to answer them in the context of the pilot study):

---

[5] Ibid.

[6] The West Midlands was chosen following an analysis of relevant crime and hospital admissions data. A 19th route was added during the life of the pilot.

[7] Chaperones received training in Mentors in Violence Prevention, trauma-informed practice, local safeguarding processes and exploitation risks. They were equipped to signpost the children and young people to a range of different types of support and were trained to build relationships with community assets such as local shops. They were made aware of the location of life-saving (and harm-reduction) equipment such as bleed control kits. The presence of a chaperone is intended to deescalate the risk of serious violence near schools and provide diversionary opportunities. They do not play an enforcement role but have been trained to deescalate conflict and respond to safeguarding concerns. Providers were selected through a tendering process, with a panel of WMVRP partners (including education reps) in each geography reviewing the readiness of applicants to deliver the project safely and in a geographically and culturally competent way.

a. Is there a significant difference in crime and/or ASB along chaperoned routes compared to the controls?
b. Is there any evidence of the spatial spillover (displacement) of crime/harm to areas adjoining the routes?
c. Do schools with chaperones show decreased student absences (and therefore higher attendance rates) and higher student attainment compared to the control schools?
d. Do schools with chaperones show increased feelings of safety among pupils, teachers and parents compared to the control schools?

Ethical review

The London School of Economics and Political Science (LSE) has already gone through the formal ethics process internally within the school.[8] Approval was received prior to the student survey rollout.

Data protection

The Centre for Economic Performance (CEP) has instituted a secure data lab for crime data that fully complies with LSE Research Laboratory Security Standards for Sensitive Data, which are publicly available on the LSE website. Key contacts and leadership roles are listed on this page as well.

Data on the secure server are stored within password-protected folders that can be accessed only by licensed users (and LSE IT staff). These folders are assigned by project, and access is granted according to licensing conditions and on approval by a centre manager of LSE.

At the end of a data contract, the folder(s) containing the contract data and any other derived files are deleted from the server. At the end of the hardware's usable life, all storage drives that have held these data are removed and degaussed using a National Cyber Security Centre–approved device before being sent for disposal.

The nature of the project makes it necessary for us to process sensitive data (E.g recorded absences and student grades); however, we will not be using student names or personally identifiable information in any part of the analysis.

This work contains statistical data from the Office for National Statistics (ONS), which is under Crown Copyright. The use of ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets that may not exactly reproduce ONS aggregates.

We acknowledge that the analysis for the education part of this report was carried out in the Secure Research Service, part of the ONS. This work was undertaken in the ONS Secure Research Service using data from the ONS and other owners; this does not imply the endorsement of either the ONS or the other data owners.

[8] The ethics approval was received by the team from the Research Ethics Committee at the London School of Economics under application number #41466. The key point of contact for this process was Lyn Grove (l.grove@lse.ac.uk).

Project team/stakeholders

Evaluation team

- Prof. Tom Kirchmaier – project director (LSE). As principal investigator Tom will be the YEF point of contact. He will coordinate the team, lead the research activities and have overall responsibility for the quality of the project outputs, the management of risks and the implementation of risk mitigation measures.
- Ms Shubhangi Agrawal – project coordinator (LSE). Shubhangi will be the contact point for the impact evaluation and will coordinate the empirical side of the project.
- Ms Julia Friedberg – policy officer (LSE). Julia will be the contact point for the impact evaluation and will coordinate the empirical side of the project.
- Ms Yasaman Saeidi – senior researcher (LSE). Yasaman will assist in conducting the data analysis.
- Mr Ricky Wang – junior researcher (LSE). Ricky will assist in conducting the data analysis.
- Mr Matteo Sandi from the LSE may also provide ad hoc as a researcher, as needed. We would also like to disclose that he is an expert with the YEF.

Methods

Participant selection

Selection of schools

- The participating schools were selected by the WMVRP before the LSE was commissioned to conduct the impact evaluation. There was a total of 24 schools in the programme.

Selection of routes

- WMVRP identified the routes in collaboration with the participating schools, local policing teams and community safety partnerships.
- The zones within each local authority where routes were placed were identified using the best available data: violence data from the mostly pre-pandemic academic year (2019–2020) and local knowledge from police officers, schools and WMVRP teams. The WMVRP strategic needs assessment and schools risk matrix were also used.
- More detailed and more current data then became available and influenced the proposed route. Routes varied in length and included streets, roads and parks. The WMVRP team carried out walkthroughs with community members and partners to identify the presence of alternative capable guardians (for example, business improvement district security staff in city centres and staff from other detached youth programmes) and areas of higher risk. The routes were then agreed between the WMVRP, providers, police and schools based on a review of this material and observations of the students' natural routes.[9]
- Local police, providers, schools, the WMVRP and local community stakeholders were all consulted, and this combined intelligence was used to maximise the reach and positive outcomes of the routes.
- The first route went live on 5 October 2021. Sites/providers worked with the WMVRP and local schools to develop routes, plan for the project's implementation and build trust with key stakeholders and the young people.[10]

Selection of chaperones

- Chaperones were employed by local service providers/sites. These were trusted youth workers placed on the routes to and from the relevant schools. Six chaperones on average were placed on each route.
- Tactical information, such as the locations for chaperones on these routes, emerged from the co-design process. Such decisions required both input from the participating schools and providers 'soft intelligence'[11] relating to the areas in which the young people might congregate and where they felt safer.

[9] The initial proposal for the routes was identified as per the above, with adjustments made following the development stage of walk-throughs and scoping. All adjustments were agreed prior to the routes' commencement; these alterations were agreed to maximise the chaperones' engagement with the young people (i.e. by situating them on the routes that the young people actually used to walk to and from school, where there was a high level of footfall).

[10] See Table A.4 in the appendix for information on the route names, the local authority and the date when the routes went live.

[11] The providers were the commissioned youth worker organisations that successfully tendered to deliver the activity. They had worked in the localities in question on various other projects with young people and so had an understanding of what was happening on those routes and were familiar with local residents ('soft intelligence'), making them best able to successfully deliver the activity. Where possible, providers based in the locality (who would have the best working knowledge and an active footprint across the area) were chosen.

Data collection

Table 1: Methods overview

Research methods	Data collection methods	Participants/data sources	Data analysis method	Research questions addressed
Quasi-experimental	Police administrative data	West Midlands Police	Differences-in-differences	RQ1, RQ2, RQ3, RQ4
Quasi-experimental	Department for Education Department of Work and Pensions Higher Education Statistics Authority HM Revenue & Customs National Pupil Database (bespoke dataset extract) Office for National Statistics dataset	West Midlands Schools, Department for Education (DfE)	Differences-in-differences Two-way fixed effects regression Matching on observables	RQ1, RQ2, RQ5
Quasi-experimental	DfE Department of Work and Pensions Higher Education Statistics Authority HM Revenue & Customs National Pupil Database (bespoke dataset extract) Office for National Statistics dataset	West Midlands Schools DfE	Differences-in-differences Two-way fixed effects regression Matching on observables	RQ1, RQ2, RQ5

Crime and antisocial behaviour data

Data on incidents of ASB and violent and property crimes was provided by West Midlands Police (WMP) in the form of extracts from police recording systems. This data covers the periods September 2014 to July 2019 and September 2021 to July 2022. This gave us five academic years of pre-treatment data. We excluded the highly anomalous COVID-19 academic years, since outcomes during the pandemic cannot be compared fairly to those before or after. To construct the sample used in our analysis, we selected only those crimes/incidents that occurred during school time.[12] 'School time' is defined as term time (September to July), with periods restricted to weekdays and daytime (06:30–17:30).[13] This is the ideal period when Step Together patrols could have an effect on crime/ASB and is in line with the evaluation of the Safe Passage programme in Chicago (McMillen, Sarmiento-Barbieri & Singh, 2019). Domestic crimes are filtered out, as Step Together is not intended to impact these. We then aggregated the data to the level of cells (see the Analysis section) and years/months. Since the number of crimes and incidents within cells were usually low, setting the unit of time for our analysis to months reduced the number of zeros in the crime/incident data.

The final sample used in our analysis was a panel of monthly ASB incidents and violent and property crime counts for each cell within the WMP Area. However, the process of receiving data from WMP took longer than expected. One cause of delay was administrative issues regarding the legal agreements for data sharing between LSE and WMP. Moreover, the data that was initially provided was scrutinised before analysis, leading to the identification of several discrepancies. This required the evaluators to ask for the data again. The data was shared by the provider upon agreement with the evaluator, which required approvals and a specific framework. Changes in the framework by the provider created delays and required the evaluators to go over the process again. This entire process took approximately five to six months. Another major cause of concern was the inconsistencies in the original data that were detected during the initial analysis. The data was recollected for assessment at the start of 2019. However, the computer system used to record crimes at WMP changed, leading to some discrepancies that we were unfortunately unable to correct. At this point, we expected that getting new data without these discrepancies would be infeasible (or, if possible, would result in more lengthy delays). Therefore, we decided to proceed with the analysis on a best-effort basis, using the data we had received. These issues may potentially affect the reliability of our results.

For future evaluations involving police administrative data, WMP recommends that evaluators provide a clear specification for data required, including specific fields, definitions, time frames and any legal agreements needed. It was possible to map crime data to routes; however, there were issues with doing this within the time frames set out. These were because of challenges in resolving data-sharing agreements and discrepancies and resulted from changes in the framework and computer system used to record crime during the course of the evaluation.

Educational data

We use the list of treated schools provided to us by the YEF. Also, as the number of these schools is limited, and there might be spillover effects on other schools in the affected areas for further analysis, we merge our school datasets with the geographical cells dataset that we used for the crime analysis. This allows us to flag the second set of treated schools based on which geographical cell they are located in.

Data on students' attendance records is provided by the Department for Education (DfE) in the Secure Research Service, part of the ONS. This data is for schools in the West Midlands area and covers the academic years 2014–2015 to 2018–2019 and part of academic year 2021–2022. This allows us to study five pre-

[12] A possible concern is that the times recorded for crimes/incidents may not be accurate. We did not notice any problems in the data to suggest this may be the case, but we cannot eliminate this possibility.

[13] As a robustness check, we also restrict the data to the starts and ends of school days, when patrols are expected to take place. Our main Poisson regression results are largely similar; however, the confidence intervals are much larger. It is unclear whether this is because of statistical error or because crime/ASB in the middle of the school day might be unaffected by Step Together patrols. Future evaluations with more data may wish to test these explanations.

treatment periods and one academic year after the treatment.[14] We have students' attendance data only for the autumn and spring terms (i.e. overall attendance data for two terms) for the above academic years. As the attendance data in the summer term of academic year 2021–2022 has not been provided to us yet, we focus only on the absence rates in the autumn and spring terms and the overall absence rates for the two terms.

For the years for which we have access to the full dataset, we observe that absence rates within a given school during the summer term are proportional to those in other terms (although they are higher in the summer, the pattern remains consistent across schools). Therefore, we do not expect that limiting our analysis to the first two terms will pose any issues. In this sense, using the outcome variable *overall two-term absence rate* will be enough for the purpose of our analysis, and the drawback of not having the data for the summer term 2022 should not be a source of concern. We define the absence rate as the percentage of all possible sessions in the specific period that the student was absent. We define the average absence rate for each school as the average absence rate for its students in the abovementioned time periods. Furthermore, as we have also been given access to absence data for academic year 2013–2014, we can utilise this additional year as a pre-trend for the first part of the analysis to gain a more inclusive pre-trend. Our results are robust to the inclusion of this additional pre-trend year.

We also want to study the students' GCSE performance in secondary schools in this area. We have data on whether individual students attempted their Maths and English GCSEs or not. We also have the highest standardised points for Maths and English GCSEs for these students. The data covers academic years 2014–2015 to 2018–2019 and the academic year 2021–2022. We calculate the percentage of all students who attempted their Maths and English GCSEs for each school. This is determined by dividing the number of students who took the GCSE English exam and the number of students who took the GCSE Maths exam in a given academic year by the total number of students in that school for the same year.

We also calculate the average score that students achieved for each one of these modules. Since 2020, GCSE scores have been given as 1 to 9, with 9 being the highest and 1 being the lowest. Pupils need a 4 for a 'standard pass' and a 5 for a 'strong pass'. Grade boundaries are decided by examiners each year (BBC News, 2024; Jadhav, 2018). Thus, we have two main secondary outcomes from the education data: a) absence rate and b) GCSE attainment (including the attempt rate for GCSE English, attempt rate for GCSE Maths, average score for GCSE English and average score for GCSE Maths).

Finally, we have to mention that we did not have access to attainment or attendance data in academic year 2022–2023, which would be the ideal post-treatment academic year. As a result, the total absence rate for that academic year is unavailable. It would be advisable to incorporate the data from this year to achieve a more precise future study.

Furthermore, for the second part of our analysis, we also utilise the provided datasets on students and school characteristics to construct a set of observable variables that can later be used in matching techniques. These variables span a wide range of characteristics that can help us to use matching techniques to find suitable control schools for the list of treated schools provided for us by the YEF. A subset of these variables consists of the percentage of students currently eligible for free meals, the percentage of students

[14] We did not have access to data from academic year 2022–23 for this study. We had access to data for the spring term of 2021–2022. Future studies could incorporate a longer post-treatment period.

who have used free meals in the last six years, the percentage of students for whom English is not their first language, the students' ethnic backgrounds, the percentage of female students and the type of educational establishment. We also utilise a dataset on the characteristics of the local authority in which the school is located. The variables we use to describe the local authority characteristics are income, employment, education, skills and training, health deprivation, crime score, housing and living environment.

We want to emphasise why school-level data, as opposed to pupil-level data, is the more suitable choice for studying educational outcomes in this context. First, the intervention is a school-wide intervention aimed at impacting the broader school environment rather than at targeting individual students directly. Since the intervention affects school policies, culture and practices collectively, the school level is an appropriate unit with which to capture these broader changes in absence rates and attainment. Also, a school-level analysis allows us to observe the aggregate changes that the intervention may bring to overall absence rates and attainment scores, which could reveal patterns that would not easily be visible at the individual level. The goal is to understand whether there is an observable difference in average outcomes for schools with and without chaperones; school-level data provides a clearer measure of these differences. Student-level data often has considerable variability due to numerous individual factors (such as socioeconomic backgrounds, prior performance or personal challenges) that may confound the results. Analysing outcomes at the school level helps control for these individual variations, allowing us to isolate the potential effects of chaperones as a structural feature of the school environment.

Finally, to capture the effect on individual students, we would need access to data with longer pre- and post-treatment periods to develop a valid counterfactual. This counterfactual could, in theory, be constructed by focusing on 'movers' – students who were initially in an untreated school and later moved to a treated school, or vice versa. However, this approach is not feasible with the dataset we currently have, as it includes less than a year of post-treatment data. To obtain valid results, it would be necessary to track students for multiple years following the intervention. Therefore, in our case, aggregating outcomes at the school level is less prone to noise and provides a more stable measure of the intervention's impact.

Analysis

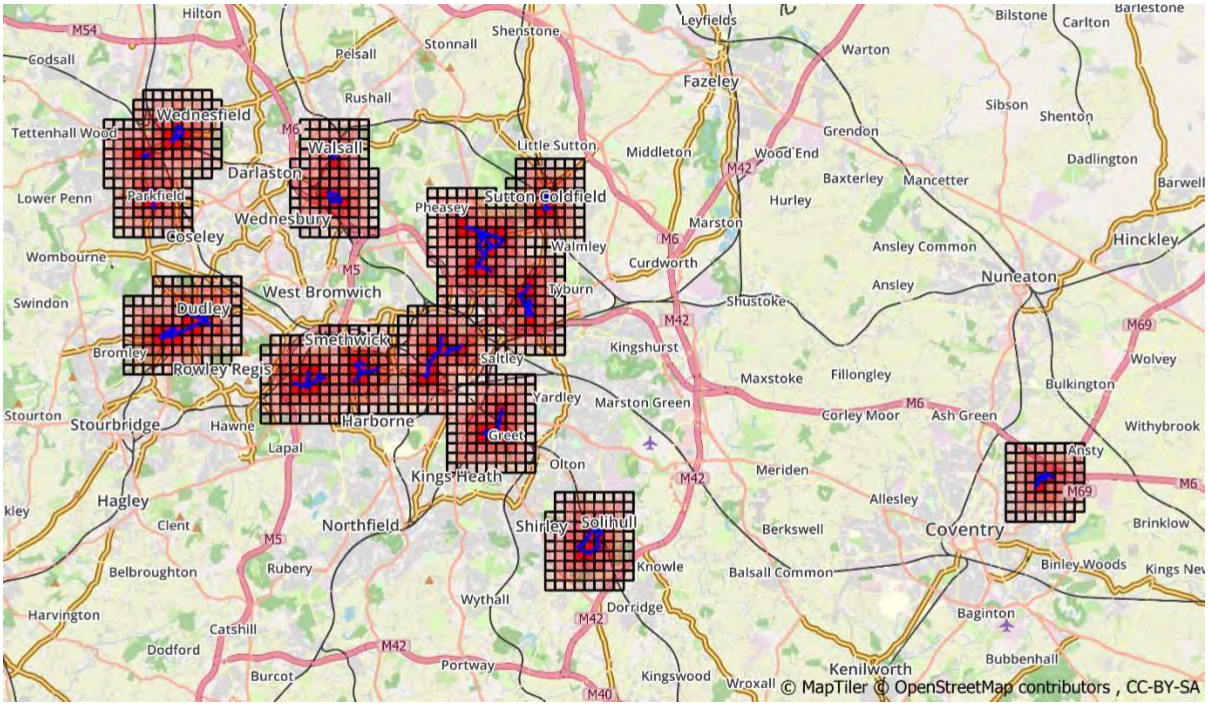
RQ2. How can a plausible counterfactual be constructed?

To measure the impact of Step Together, we adapt a method used successfully in the evaluation of Safe Passage in Chicago (McMillen et al., 2019). This method is also in line with the study plan. First, a grid of 500 m × 500 m cells is constructed over the WMP area. These cells are defined using Ordnance Survey National Grid eastings and northings, with new cells starting at multiples of 500 m. A map of cells and Step Together routes can be found in Figure 1. Cells that overlap with a route are then marked, forming the treatment group. Cells up to three squares adjacent to the routes are also marked and will serve as the control group. This helps to mitigate bias due to unbalanced observable and unobservable characteristics, under the assumption that the areas immediately adjacent to the route cells are similar to the route cells themselves. We chose this method under the plausible assumption that nearby areas have similar characteristics.

The Safe Passage programme in Chicago is similar to Step Together, justifying our adoption of the method used by McMillen et al. (2019). A potential concern, however, is that geographical differences between Chicago and the West Midlands may translate to neighbouring cells not serving as good counterfactuals in the West Midlands. In particular, crime in England tends to be concentrated around high streets and other hotspots, while in Chicago, crime is likely to be more evenly distributed due to the grid-like structure of the city's streets. We attempt to address this concern by testing smaller cells of 100 m × 100 m, all of which yielded similar regression results.[15]

The ideal (but an infeasible) method of constructing plausible counterfactuals would be to run a randomised controlled trial (RCT) for Step Together, under which counterfactuals would be readily available. An alternative method of constructing a counterfactual, analogous to controls from an RCT, would be to construct 'synthetic' routes around other schools in the West Midlands that have not implemented the Step Together programme. Some form of matching would be needed to ensure that the synthetic routes are similar to the actual treated routes. It is likely that this would necessitate a large amount of data – potentially offering more detail than the data that was available to us – to generate high-quality matches and, hence, plausible counterfactuals.

Figure 1: Map of the routes and cells



Notes: This map shows Step Together routes (in blue) and cells constructed as described above. Cells closest to routes are brightest, with cells further away fading in intensity. Only cells at most three squares away from on-route cells are shown – these are all treatment and control cells.

Analysis of crime and antisocial behaviour

This section details the analysis undertaken to answer RQ3 and RQ4.

After the grid cells are constructed (as previously described), crime and ASB incidents are also mapped onto the same grid cells using the location data provided for each crime/incident. This allows us to compare ASB and crime outcomes pre- and post-treatment using a difference-in-differences (DiD) approach. By including time and cell fixed effects, the DiD approach controls for unobserved heterogeneity and temporal variations

[15] One issue with smaller cells is that the number of zero-offence observations is inflated since most cells already have fairly low offence counts. Future evaluations that consider smaller cells should bear this in mind.

that may affect the dependent variable, including possible seasonality.

Our baseline regression specification is:

$$y_{it} = \beta_1 OnRouteCell_i \times Post_t + \beta_2 OneCellAway_i \times Post_t + \beta_3 TwoCellsAway_i \times Post_t + \gamma_i + \delta_t + e_{it}, \quad (1)$$

where:

- y_{it} is the outcome variable: the number of crimes or ASB incidents in cell *i* in month-year *m*.
- $Post_t$ is an indicator variable that has the value 1 in the post-treatment period (the 2021–2022 academic year).
- $OnRouteCell_i$ is an indicator variable for cell *i* being on top of a route
- $OneCellAway_i$ and $TwoCellsAway_i$ are indicator variables for cell *i* being one or two cells away from an on-route cell, respectively.
- γ_i and δ_t are cell and month-year fixed effects, respectively.
- e_{it} is the error term.

Since the outcome variables consist of count data, our preferred specification uses a Poisson regression to estimate the coefficients. A Poisson regression models the log-transformed expected count as a linear function of the predictor variables, which means that coefficient estimates should be interpreted (approximately) as percentage changes and not as changes in levels. Poisson regressions are also used when there are a large number of zeros in the data.

In order to isolate the causal effect of Step Together patrols on crime and ASB, it is crucial to assume that the outcomes in the treated cells and nearby non-treated cells would have evolved similarly in the absence of treatment. This is known as the parallel trends assumption. If the assumption that nearby cells are similar to on-route cells is true, then the parallel trends assumption will also be true, since crime and ASB would evolve in parallel over time in the absence of treatment.

To further examine any changes in the treatment effect over time, we also run dynamic DiD regressions, allowing for a different effect in each year-month and thereby accounting for seasonality. This also enables the parallel trends assumption to be tested somewhat, as it allows for the examination of pre-trends. However, breaking down the treatment effect into more granular monthly treatment effects results in larger standard errors, since there is effectively less data available to estimate each coefficient.

The dynamic regression specification is identical to the static specification, except the first term is allowed to vary over time:

$$y_{it} = \sum \beta_{1t} OnRouteCell_i \times I(t = s) + \beta_2 OneCellAway_i \times Post_{ts} + \beta_3 TwoCellsAway_i \times Post_t + \gamma_i + \delta_t + e_{it}, \quad (2)$$

where *s* indexes all year-months from the 2014–2015 to 2021–2022 academic years.

As previously described, a limitation of the data is the presence of discrepancies caused by a change in computer systems at WMP. During the period of analysis, WMP switched from one data recording system to another, and hence, the crime data originates from two separate extracts. While conducting our analysis, we found some anomalies in the number of crimes reported around 2019, when the WMP data recording system changed. It has unfortunately not been possible to fully resolve this issue; however, the inclusion of year-month fixed effects in our regression specifications will reduce any effects it might have.

Analysis on educational outcomes

In the second part of our analysis on educational outcomes, we first follow the exact same steps as above. We have geolocated each school in the grid of 500 m × 500 m cells that we have constructed over the West Midlands. This subset of schools will include both the participating schools and those schools that are non-participating but are located in the same cell as a route. Depending on the status of the cells that each school is located in, each school will be flagged either as being on a treated route, one cell away from a treated route, two cells away, three cells away or more than three cells away.

Similarly, our baseline regression specification is:

$$y_{it} = \beta_1 OnRouteCell_i \times Post_t + \beta_2 OneCellAway_i \times Post_t + \beta_3 TwoCellsAway_i \times Post_t + \gamma_i + \delta_t + e_{it}, \quad (3)$$

where:

- y_{it} is the outcome variable of school *i* in academic year *m.*
    - In the first regression, the outcome variable is the average absence rate in the school.
    - In the second regression, the outcome variable is the average standardised score in GCSE Maths.
    - In the third regression, the outcome variable is the average standardised score in GCSE English.
- $Post_t$ is an indicator variable that has a value of 1 in the post-treatment period (the 2021–2022 academic year).
- $OnRouteCell_i$ is an indicator variable for school *i* being on a cell that is on top of an on-route (a treated route).
- $OneCellAway_i$ and $TwoCellsAway_i$ are indicator variables for school *i* being on a cell that is one or two cells away from an on-route cell, respectively.
- γ_i and δ_t are school and academic year fixed effects, respectively.
- e_{it} is the error term.

This allows us to compare the absence rate and the GCSE Maths and English scores pre- and post-treatment using a DiD approach. By including time and school fixed effects, the DiD approach controls for unobserved heterogeneity between schools and for temporal variations that may affect the dependent variable.

There may also be spillovers between the schools located on treated cells and schools located on adjacent ones. To check if the results that we have found are robust, we eliminate the cells and the schools wherever they are one, two or three cells away. We only keep the schools that are in the cells, with the treated routes as the treatment group, and we mark the schools that are more than three cells away from an on-route cell as the control group. Then, we run the above regressions using only this new subset of the dataset. This will alleviate the possibility of spillovers but weaken the validity of our identification. To capture the causal effect of Step Together patrols on absence rate and GCSE attainment, it is crucial to assume that the outcomes in the treated and control schools would have evolved similarly in the absence of treatment. This is the parallel trends assumption.

One way to examine for parallel trends is running an event study and following the changes in the treatment effect over time, thus allowing for a different effect in each academic year. However, breaking down the treatment effect into yearly treatment effects results in larger standard errors, since there is effectively less data available to estimate each coefficient, and it is not necessarily the best way to prove that any pre-trends were not existing. We also visualise the trend before and after the treatment to check for the trends.

We are studying a specific limited geographical region, and we are only focusing on proximate schools, coupling this with the incorporation of school fixed effects to control for unobservable confounders. We also control for the academic-year fixed year. However, the academic-year 2021–2022 is after the pandemic; there might be a differential effect on absence rates in schools during the pandemic, so the assumption that this year fixed effect is the same for both treated and non-treated schools might not hold.

Matching

In the next step, we will redo our analysis using the exact treatment and control flags for participating and non-participating schools, which have been provided to us by the YEF. These flags were not previously used in the first step of the analysis due to the low number of participating schools that also appear in the provided dataset. We have, in total, 24 treated (participating) schools and 171 not-treated schools, which are the candidates for the control group. The non-treated schools are located in the West Midlands, are not participating in the YEF intervention and are available in the National Pupil Database (NPD). Our preferred approach is to use matching techniques, as they produce less noisy results and tolerate less imbalance while also better satisfying the parallel trends assumption. However, due to the limited number of participating schools and the small geographic area where the intervention was applied – leading to restricted opportunities for generating matched samples – we will also present the results from the baseline approach.

In this matching exercise, we will use different matching techniques to choose a subset of the non-participating schools as our control group. The matching techniques we use are one-to-one matching, k-nearest neighbour matching and Mahalanobis distance matching. We should mention that, given the small geographic area to which the intervention was applied and the limited number of schools in this area, there is limited scope for generating matched samples.

We use a wide range of variables on the different characteristics of the schools, their pupils and the neighbourhoods that the schools are located in to conduct the matching techniques. The outcome variables for all these schools are not available for all years, and we have missing values for some of the schools in random academic years. We use 2019 as our baseline when doing the matching, as this year has the lowest number of schools with missing outcome variables out of the untreated years.

We try different subsets of observables to find suitable matches. The main observable variables that have been chosen to find matches are: the percentage of students eligible for free meals; the percentage of students using free school meals in the last six years; the percentage of students for whom English is not their first language; the percentage of students with White, Black, Asian and Mixed ethnic backgrounds; the percentage of female students; the type of educational establishment; prior student performance (GCSE results); prior student recorded absences; and neighbourhood characteristics. This last variable considers the following variables for the local authority in which the schools are located: income, employment, education, skills and training, health deprivation, crime scores, housing and living environment.

In the first step, we use one-to-one matching to choose the control group. Then, we use k-nearest neighbour matching, choosing k = 3 to find suitable matches. Finally, we use Mahalanobis distance matching, which matches observations based on the Mahalanobis distance to reduce bias.

After choosing the proper control group using the aforementioned matching methods, we run the following regression to assess the effect of participating in the programme on the students' performance and absence records:

$$yit = \beta 1 Treatedi \times Postt + \gamma i + \delta t + eit, \quad (4)$$

where:

- yit is the outcome variable of school *i* in academic year *m.*
  - In the first regression, the outcome variable is the average absence rate in the school.
  - In the second regression, the outcome variable is the average standardised score in GCSE Maths.
  - In the third regression, the outcome variable is the average standardised score in GCSE English.
- $Post_t$ is an indicator variable that has a value of 1 in the post-treatment period (the 2021–2022 academic year).
- $Treatedi$ is an indicator variable showing whether school *i* has participated in the programme.
- γi and δt are the school and academic year fixed effects, respectively.
- eit is the error term.

Findings

Table 3: Participant and school details

Participants (unit of analysis)	**Treated and control units for crime analysis** For the primary crime outcomes, cells of approximately 500 m × 500 m will be created, and any cell that contains part of a Step Together route will be considered as treated. For the comparison units, we will use adjacent cells. **Treated and Control units for education analysis** Treated units are those schools that are located within cells of approximately 500 m × 500 m that contain part of a Step Together route.
Description of the settings	**Baseline characteristics of the treated and control units (crime data)** Prior to treatment and excluding 2019, property crimes in on-route cells averaged just over two per month, while the number in neighbouring cells was much lower, averaging around one per month. For violent crime, the treatment areas again have higher levels in the pre-treatment years, although the difference is not as large. There appears to be a larger increase in the average number of violent crimes per month in the on-route cells compared to that in the cells away from the routes. However, the increase appears to be roughly proportional to pre-treatment levels. Property crime appears to have fallen slightly more in on-route cells than in other cells. **Baseline characteristics of the treated and control units (education data)** **Attendance**: The absence rate in treatment areas is higher in the pre-treatment years, although the difference is not that large. The absence rate in all types of schools in the first academic year pre-treatment is between 0.04 and 0.06, which means that, on average, students attend 94 to 96 per cent of all possible sessions in an academic year. Students at schools that are located in cells containing a treated route have lower attendance rates. **Attainment**: When measured as the average standardised points in GCSE Maths, attainment in the treatment schools is lower in the pre-treatment years.
Number of participants	**Treated and control units for crime analysis** The number of treatment (on-route) cells is 104, and the number of control cells is 363. **Treated and control units for education analysis** For the absence rates and GCSE attainment outcomes (the first methodology), there are 112 treated schools. We have 98 schools in our control group. There are 35 schools in cells that are one cell away from a cell containing a treated route and 63 schools in cells that are two cells away from a cell containing a treated route. There are also 532 schools that are in cells more than three cells away from a cell containing a treated route. We use these schools as control schools in the second part of this exercise. For the second methodology, we use the participating schools as the treated group and use matching techniques to choose the control group from non-participating schools. We have, in total, 24 participating and 171 non-participating schools, all of which are candidates for the control group. These numbers may vary, since the outcomes for some of the schools are missing from the datasets provided for some of the years. They may also differ depending on the matching method used.

Evaluation feasibility

Evidence of promise

RQ1: Can we collect the necessary crime and education data within the initial time frame set out for the evaluation?

Due to the challenges we faced during the data collection process, we were unable to collect the crime and education data within the initial time frame set out. For the crime data, these challenges relate to the previously mentioned delays in receiving data from WMP. For future evaluations involving police administrative data, WMP recommends that evaluators provide a clear specification of the data required, including specific fields, definitions, time frames and any legal agreements needed. With these recommendations in mind, future evaluations may be able to collect the necessary crime data in a shorter time frame.

The education-related outcome dataset was collected from the DfE. The data access and the analysis were carried out in the Secure Research Service, part of the ONS. The data provision, the provision of a secure environment to carry out the analysis, and the transfer of external shapefiles needed for analysis took up substantial time (about three to four months). Moreover, any result to be published using NPD data needed to be reviewed by the ONS (which took up approximately two months). This time needs to be accounted for in advance, as it delays the process of analysis. The accuracy of the data provided by the DfE will affect the reliability of our result.

Additionally, the evaluators rolled out a student survey to get a baseline estimate of student feelings of safety on school routes due to the presence of chaperones. The survey was rolled out in approximately 30 schools through the Qualtrics survey platform. However, response rates were extremely low (below 10 per cent), and as a result, the evaluators decided to discontinue the survey. At this point, we do not have a sense of what type of information the teachers received, how much they encouraged students to complete the survey and how much time they granted students to complete the survey. For future studies, we suggest that it might make more sense to print physical copies and have someone present in the schools to administer them.

Future studies also need to carefully consider the time that might potentially be taken up by legal agreements, admin work and data inspection and sharing before setting up the timeline for analysis. Moreover, the involvement of different parties in data provision and in reviewing the analysis can substantially affect the timeline for the evaluators. Thus, a carefully constructed timeline that accounts for all these details is needed before an initial time frame is agreed. Ideally, a time period of two-and-a-half years is required for data collection, analysis, evaluation and reporting.

RQ2: How can a plausible counterfactual be constructed?

For violent and property crime, neighbouring cells appear to serve as plausible counterfactuals for on-route crime. While it is never possible to directly test the parallel trends assumption, the fact that our estimates in the pre-treatment period of the dynamic DiD regression (shown in Figure 3) are insignificant provides some reassuring evidence in favour of the parallel trends assumption.

On the other hand, for ASB, some pre-treatment estimates (shown in Figure 5) are meaningfully different from zero, suggesting that the parallel trends assumption may have been violated. This does not contradict the plausibility of the counterfactual for violent and property crime, as ASB is likely to have different drivers than violent and property crime and hence may be distributed differently.

A further complication is the COVID-19 pandemic. Since crime and ASB were extremely anomalous during the pandemic, we excluded data from the 2019–2020 and 2020–2021 academic years. Implicitly, this makes the assumption that crime and ASB returned to pre-pandemic patterns in 2021–2022. It is possible, however, that the pandemic structurally changed crime and ASB patterns, resulting in treatment and control areas diverging even without treatment. Note that our parallel pre-trends analysis excluded two years of crime data corresponding to the COVID-19 pandemic period of 2020–2021, when crime rates were unusual. Even if the neighbouring cells were a plausible counterfactual pre-pandemic (i.e. 2016–2019), it is unclear whether they would remain plausible counterfactuals post-pandemic. Future evaluations are likely to be less impacted by the pandemic, since they may be able to consider only post-pandemic data.

For educational outcomes, our preferred approach uses matching techniques, applying the exact treatment and control flags for participating and non-participating schools provided by the YEF. We have 24 treated (participating) schools and 171 non-treated schools as candidates for the control group. The non-treated schools are those located in the West Midlands that did not participate in the YEF intervention and are available in the NPD. The counterfactual is constructed using various matching techniques, including one-to-one matching, k-nearest neighbour (k = 3) matching and Mahalanobis distance matching.

To create robust counterfactuals, we use a wide range of variables, reflecting school, pupil and neighbourhood characteristics. Key matching variables include: the percentage of students eligible for free meals; the percentage of students using free school meals in the last six years; the percentage of students for whom English is not their first language; the percentage of students with White, Black, Asian and Mixed ethnic backgrounds; the percentage of female students; the type of educational establishment; prior student performance (GCSE results); and prior student recorded absences. Additionally, we consider neighbourhood characteristics, including local authority-level variables such as income, employment, education, skills and training, health deprivation, crime scores, housing and living environment.

We conducted an additional analysis to include a larger number of schools from the region in our study. While this is not our preferred method, it allows us to explore a broader range of potentially affected schools. In this approach, we first geolocated each school within a grid of 500 m × 500 m square cells situated over the West Midlands. This subset includes both participating schools and non-participating schools that are located in the same cell as a route. Based on the cell each school occupies, schools are flagged as being on a treated route or one, two, three or more than three cells away. The treated schools are those located on a cell with a treated route, while the counterfactuals are schools located two, three, or more cells away from a treated route.

RQ3: Is there a meaningful difference in crime and/or antisocial behaviour along chaperoned routes compared to that along controls?

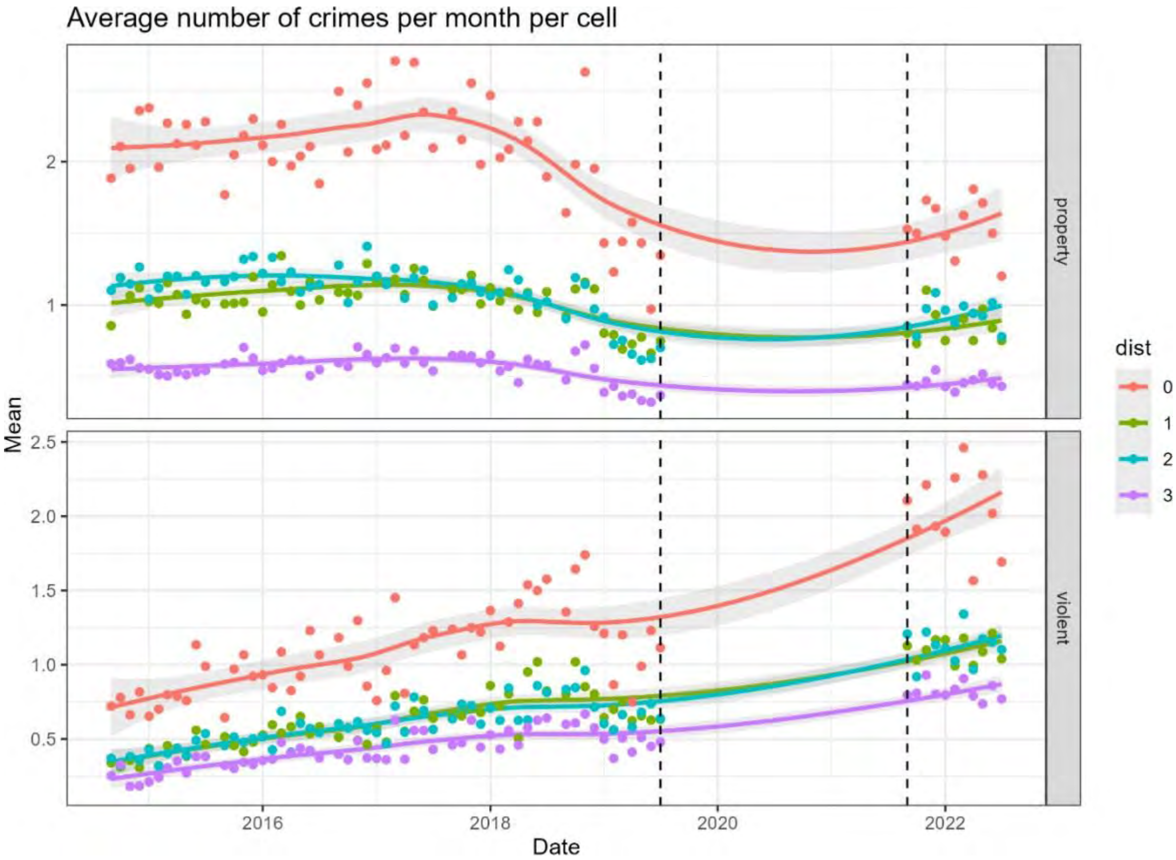
Violent and property crime

Figure 2 shows the average number of non-domestic violent and property crimes per cell each month. Prior to treatment and excluding 2019, property crimes in on-route cells averaged just over two per month, while those in neighbouring cells were much lower, averaging around one per month. For violent crime, treatment areas again have higher levels in the pre-treatment years, although the difference is not as large. There appears to be a larger increase in the average number of violent crimes per month in on-route cells compared to that in cells away from routes. However, the increase appears to be roughly proportional to pre-treatment levels. In comparison, property crime appears to have fallen slightly more in on-route cells than in other cells. Overall, trends for the treatment and control areas appear to diverge over time, suggesting that the parallel trends assumption would likely be violated if a linear model were to be used. Our preferred Poisson model is more suitable, as it allows for exponential trends, rather than linear.

Table 4 shows DiD regression estimates for violent and property crime, with the coefficient of interest being *On-route × Post*. Estimates using linear models are included for completeness only and should be interpreted with caution, since the parallel trends assumption is likely to be violated. Under our preferred specifications and using the Poisson regression, standard errors and, hence, confidence intervals (not shown) are large, and therefore, we are unable to conclude that crime along the routes was reduced. The point estimate for property crime is fairly large in magnitude, corresponding to approximately an 8 per cent reduction in crime, but smaller than the estimated minimum detectable effect of an 11 per cent reduction required to achieve an effective power of 0.5 (see appendix). Additional data may help improve power and provide a more conclusive result.

Further analysis using a dynamic DiD specification (as per Equation 2) reveals similar results, with estimates and 95 per cent confidence intervals presented in Figure 3. All confidence intervals in the post-treatment period cross zero, echoing the null result from the static regression.

Figure 2: Crimes pre- and post-treatment



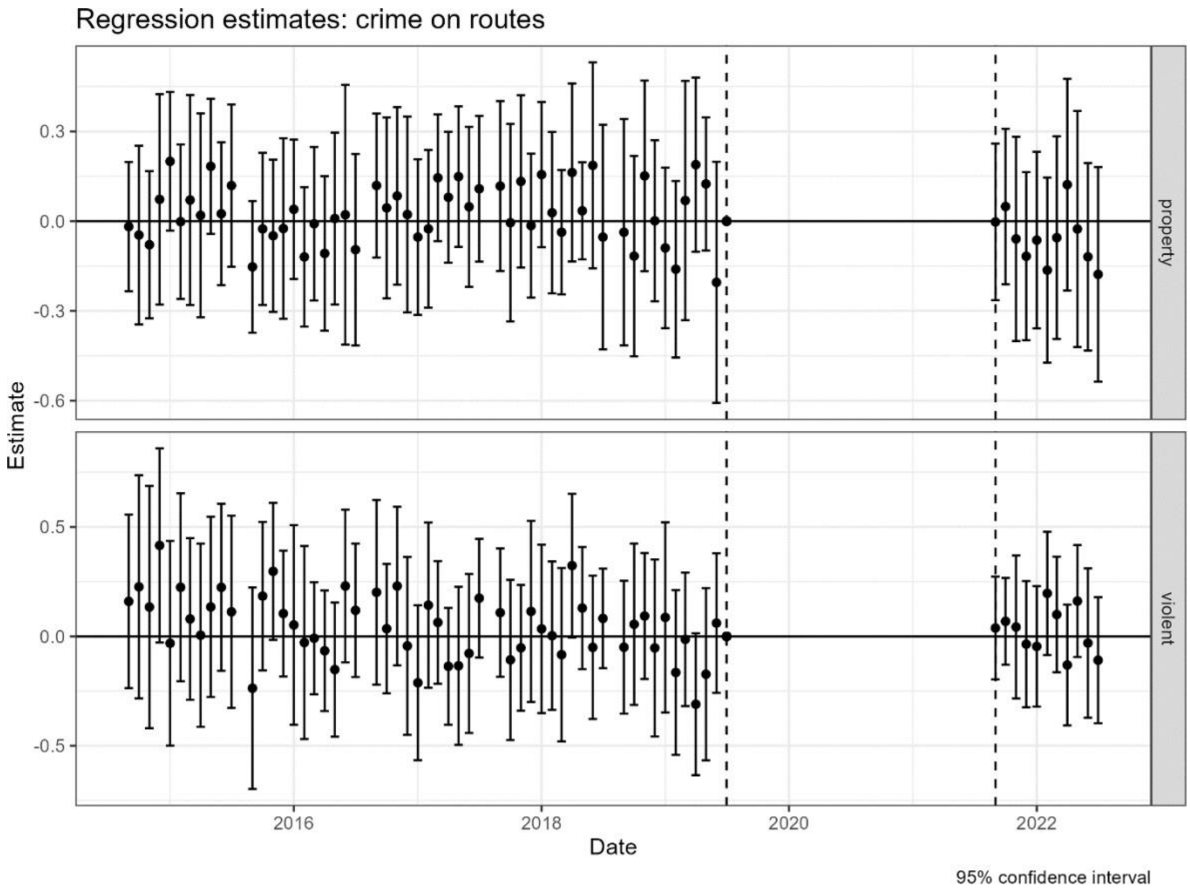
Notes: This figure shows the average number of non-domestic property and violent crimes during school times per month per cell. The variable *dist* refers to the distance of a cell to the nearest Step Together route. For example, cells where *dist* = 0 are on routes, while those where *dist* = 1 are one cell away from a route (or, equivalently, an on-route cell). The dashed vertical lines indicate the end of pre-treatment and the start of post-treatment periods, respectively.

Table 4: Difference-in-differences regression output for crimes

	Violent crime		Property crime	
	Linear	Poisson	Linear	Poisson
On-route × Post	0.5635 (0.1203)	–0.0122 (0.0677)	–0.4143 (0.2256)	–0.0834 (0.0851)
1-away × Post	0.1029 (0.0419)	–0.0508 (0.0336)	–0.0814 (0.0727)	0.0082 (0.0642)
2-away × Post	0.1415 (0.0674)	–0.0043 (0.0471)	–0.0656 (0.1368)	0.0365 (0.1008)
Cell Fixed Effect	Yes	Yes	Yes	Yes
Year-month FE	Yes	Yes	Yes	Yes
Sample size	66,132	64,878	66,132	65,802

Notes: Outcome variables are crimes per cell per month during school times. *On-route* refers to cells through which a Step Together route runs. *1-away* and *2-away* refer to cells one and two cells away, respectively, from the nearest on-route cell. All standard errors are clustered by the Step Together route.

Figure 3: Dynamic DiD estimates (crime)



Notes: This figure presents estimates from dynamic DiD regressions for property and violent crime (black dots), with 95% confidence intervals (black bars) shown. Outcome variables are crimes per cell per month during school times.

Antisocial behaviour

Figure 4 shows average incidents of ASB per cell per month. During the pre-treatment period, there is a large gap in ASB incidents for the treatment and control cells, which shrinks drastically post-pandemic, an effect possibly caused by Step Together. However, the trend for on-route cells differs from that for other cells, peaking around 2017. (Note that comparing raw numbers does not conclusively prove a violation of the parallel trends assumption.)

Table 5 presents DiD regression estimates for ASB that have been estimated using Equation 1. Here, we see large negative estimates using both linear and Poisson (preferred) models. The Poisson estimate of –0.1950 corresponds to a roughly 18 per cent reduction in ASB — a large effect size. These results should, however, be interpreted with some caution, given the potential violation of parallel trends.

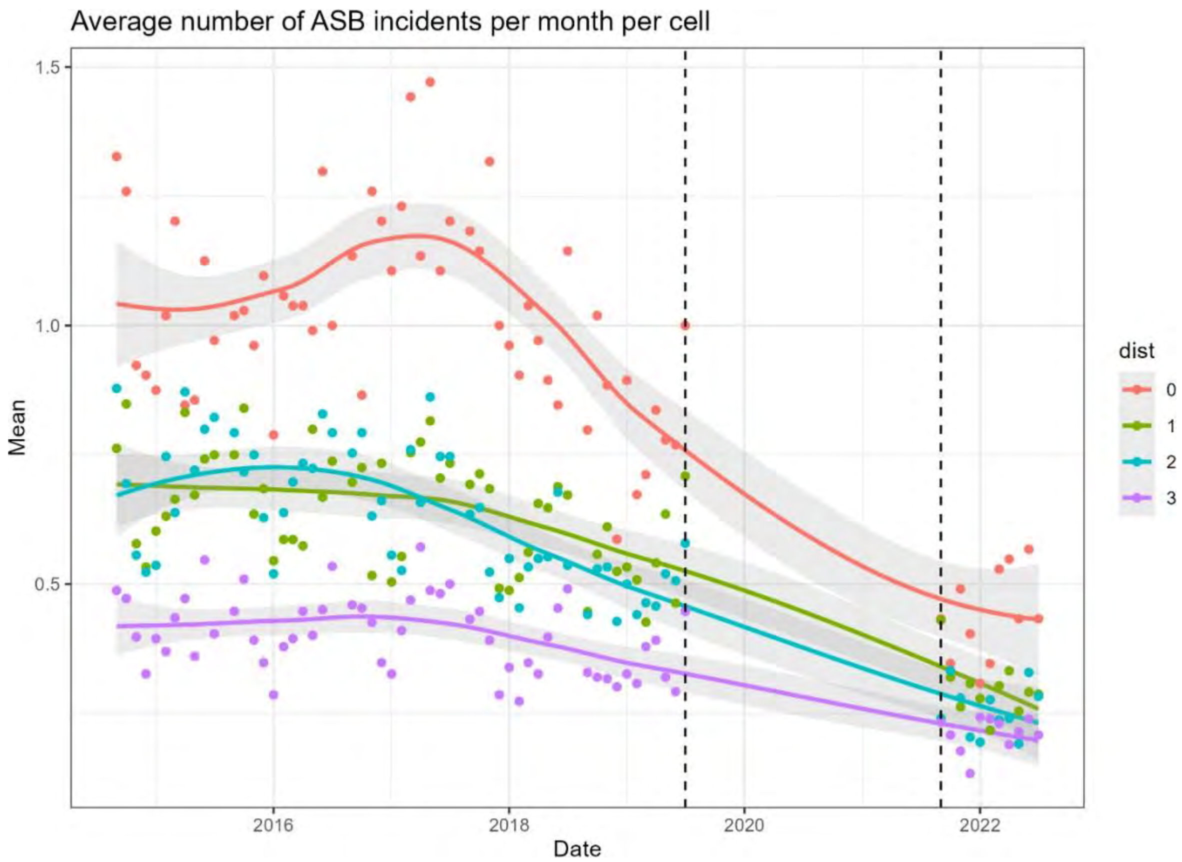
Figure 5 shows results from the dynamic DiD Poisson regression estimated using Equation 2. In the post-treatment period, most of the confidence intervals cross zero due to a lack of statistical power. The negative point estimates are concentrated in the first half of the 2021–2022 academic year, suggesting that the reduction in ASB from Step Together was likely strongest during this period of time. As with the static regression, however, some caution must be applied with respect to the validity of the results. A substantial number of the pre-treatment period estimates are non-zero, suggesting that the parallel trends assumption is likely violated.

Summary

Overall, we are unable to conclude that Step Together has had an impact on either property or violent crime. The estimates for both outcomes are close to zero, and standard errors are large due to a lack of statistical power. This lack of precision in the estimates means that it is difficult to say whether Step Together meaningfully affected crime, and while we do find a meaningful negative effect for ASB, the conditions required for the DiD method to be valid may not be met, and hence, the results may be biased.

Note that these results are not adjusted for multiple comparisons. Due to computational difficulties, we are unfortunately unable to implement the Romano–Wolf correction for multiple hypotheses (Romano & Wolf, 2005, with later improvements). Simpler, less powerful methods, such as the Holm–Bonferroni correction (Holm, 1979), are feasible. However, given that this evaluation already suffers from a lack of power, we chose not to adjust the results. Readers should bear this in mind when interpreting the results.

Figure 4: Antisocial behaviour pre- and post-treatment



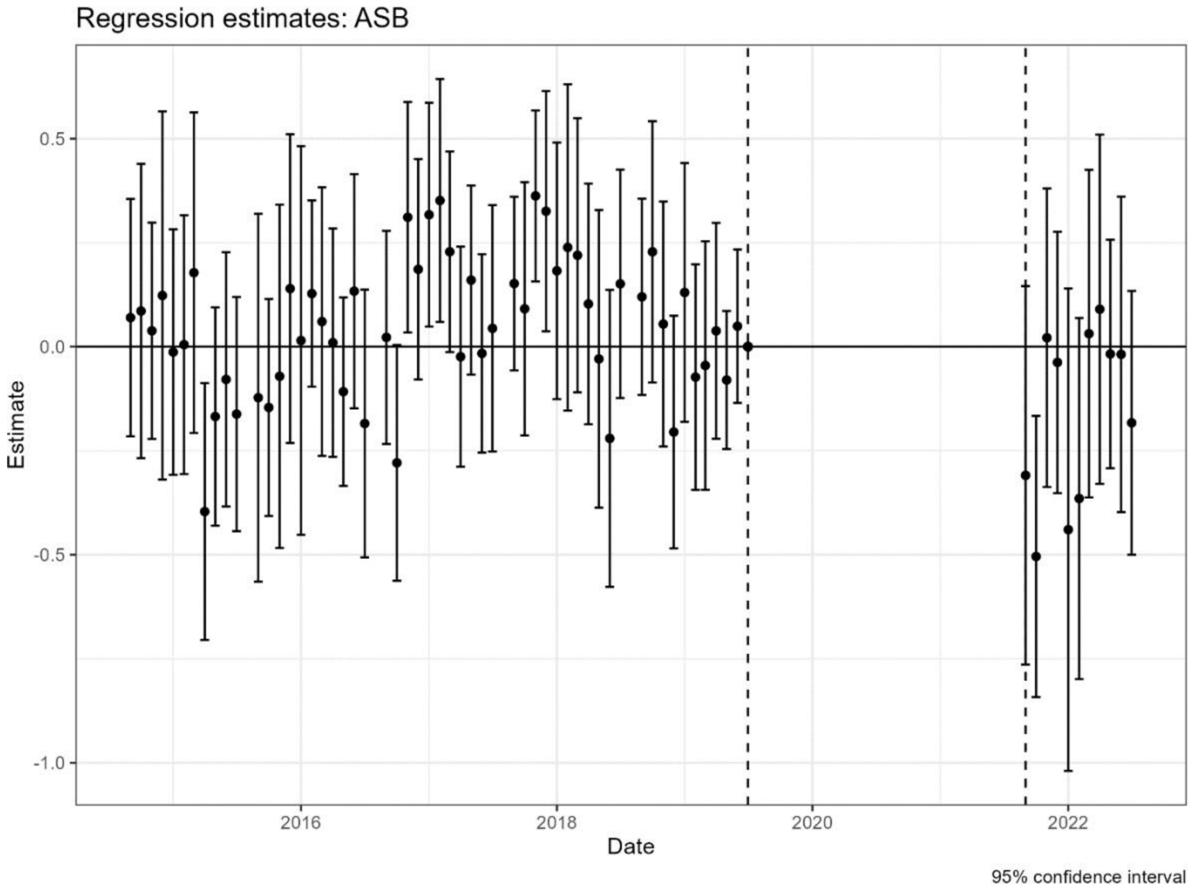
Notes: This figure shows the average number of ASB incidents during school times per month per cell. The variable *dist* refers to the distance of a cell to the nearest Step Together route. For example, cells where *dist* = 0 are on routes, while those where *dist* = 1 are one cell away from a route (or, equivalently, an on-route cell). The dashed vertical lines indicate the end of pre-treatment and the start of post-treatment periods, respectively.

Table 5: Difference-in-differences regression output for antisocial behaviour

	Antisocial behaviour	
	Linear	Poisson
On-route × Post	–0.3895 (0.0862)	–0.1950 (0.0846)
1-away × Post	–0.1519 (0.0683)	–0.1185 (0.0645)
2-away × Post	–0.1868 (0.1410)	–0.2622 (0.1427)
Cell FE	Yes	Yes
Year-Month FE	Yes	Yes
Sample size	64,284	64,284

Notes: Outcome variables are crimes per cell per month during school times. *On-route* refers to cells through which a Step Together route runs. *1-away* and *2-away* refer to cells one and two cells away, respectively, from the nearest on-route cell. All standard errors are clustered by the Step Together route.

Figure 5: Dynamic difference-in-differences estimates (antisocial behaviour)



Notes: This figure presents estimates from the dynamic DiD Poisson regression for antisocial behaviour, with 95% confidence intervals shown. Outcome variables are crimes per cell per month during school times.

RQ4: Is there any evidence of spatial and/or temporal spill over (displacement) of crime or harm to the areas adjoining routes?

Violent and property crime

The additional regression terms in Table 4 allow us to test for the potential displacement of crime from on-route cells to neighbouring cells. We do not find any evidence for such displacement effects, as seen by the near-zero estimates for *1-away × Post* and *2-away × Post*.

Antisocial behaviour

In Table 5, the large and negative coefficient estimates for *1-away × Post* and *2-away × Post* provide some evidence that ASB in the cells adjacent to the routes was reduced. However, the standard errors are fairly large due to a lack of statistical power, and the possibility that the parallel trends assumption may not hold means that these estimates should be interpreted cautiously.

RQ5: Do schools with chaperones show decreased student absences (and therefore higher attendance rates) and higher student attainment compared to that seen in control schools?

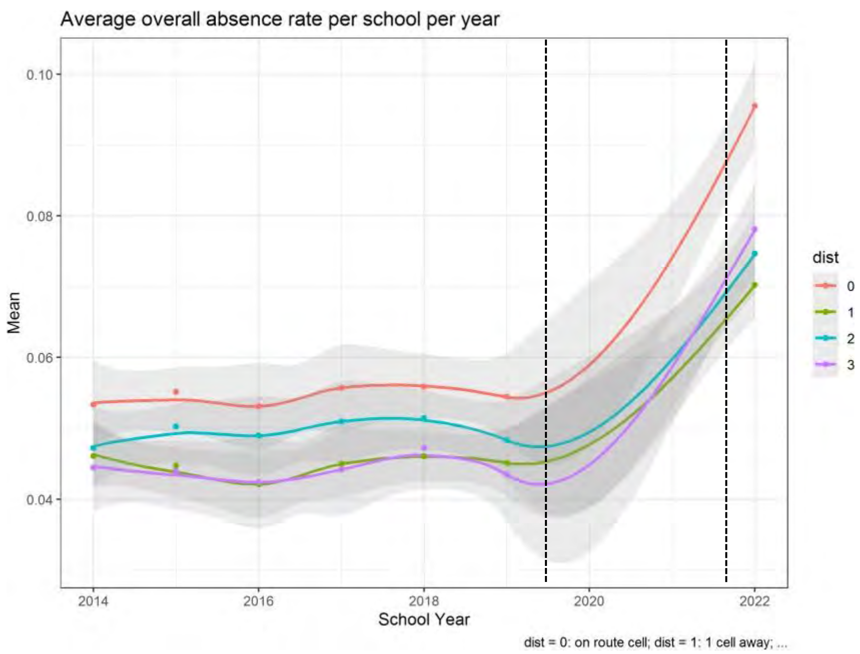
Figure 6 shows the average absence rate in schools in each type of cell (schools in cells containing a treated route or one, two or three cells away) in each academic year. The absence rate in the treatment areas is higher in the pre-treatment years, although the difference is not that large. The absence rate in all types of schools in the first academic year before treatment is between 0.04 and 0.06, which means that, on average, students attend 94 to 96 per cent of all possible sessions in an academic year. Students at schools located in cells containing a treated route have higher absence rates and, therefore, lower attendance rates.

Post-treatment, there appears to be a larger increase in the average absence rate per year in schools located in on-route cells than in school located in the cells away from the routes. In pre-treatment years, the difference between the absence rates for schools located on routes and those located one cell or two cells away are around 1 per cent, increasing to around 2 per cent after treatment.

Table 6 shows DiD regression estimates for absence rates, with the coefficient of interest being *On-route × Post*. Estimates using linear models are included for completeness only and should be interpreted with caution, since the parallel trends assumption might be violated. The ordinary least squares (OLS) model does not show any significant effect for the schools located in the treated cells. However, we observe significant estimates for *1-away × Post* and *2-away × Post*.

Therefore, there is a negative effect on absence rate that is equivalent to a positive effect on students' attendance in schools located one cell or two cells away, which could suggest a potential spillover of the Step Together programme to adjacent regions. Under the Poisson regression, neither estimate of interest is statistically significant at the 5 per cent level, and therefore, we are unable to conclude that absences decreased in schools along the routes.

Figure 6: Absence rate pre- and post-treatment



Notes: This figure shows the average absence rate in schools during autumn and spring terms of each academic year in each type of cell. Here, *dist* refers to the distance between a cell that a school is located on and the nearest Step Together route. For example, schools where *dist* = 0 are located in cells on routes, while those where *dist* = 1 are located one cell away from a route (or, equivalently, one cell away from an on-route cell)) (Source of data: ONS)..

Table 6: Difference-in-differences regression output for absence rates

	Absence rate	
	Linear	Poisson
On-route × Post	0.002 (0.002)	–0.026 (1.007)
1-away × Post	–0.013*** (0.004)	–0.136 (1.939)
2-away × Post	–0.013*** (0.003)	–0.175 (1.425)
School FE	Yes	Yes
Year FE	Yes	Yes
Sample size	1,470	1,470

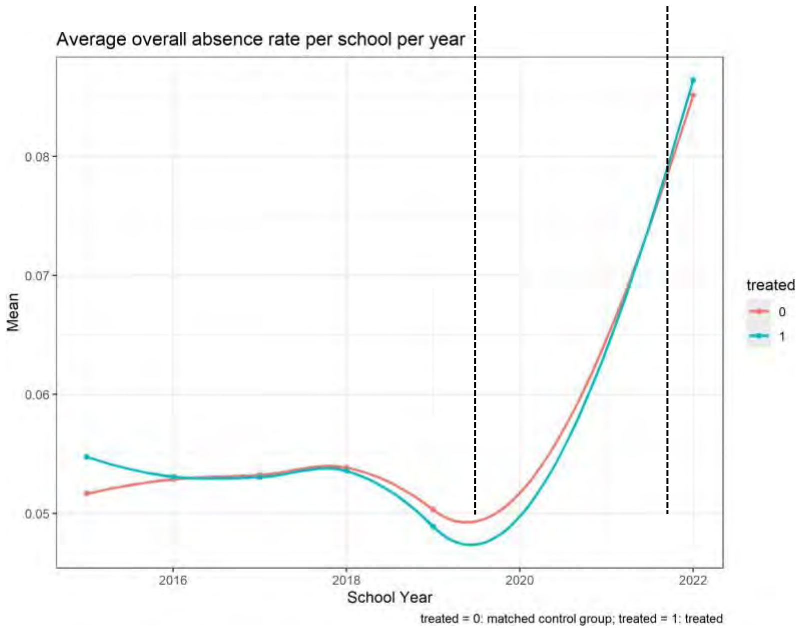
Notes: Outcome variables are the overall absence rate (absence rate in the autumn and spring terms of each academic year) in the schools in each type of cell. *On-route* refers to schools located in cells through which a Step Together route runs. *1-away* and *2-away* refer to schools in cells one and two cells away, respectively, from the nearest on-route cell. All standard errors are clustered by the Step Together route. * p < 0.10; ** p < 0.05; *** p < 0.01 (Source of data: ONS). .

We do another exercise using schools more than three cells away from treated schools as the control. The pre-trends of absence are visually parallel under this setting. We do not find any significant effect on school absences in the treated or control schools after treatment in this setting. The related figure (Figure A.2) and the coefficient can be found in the appendix.

In the next step, we provide the results of the regressions, using three different matching techniques to select the control groups. Figures 7, 8 and 9 show the average absence rate in the treatment and control schools in each academic year. The treatment schools are the participating schools, and the control schools are selected using one-to-one, k-nearest neighbour (k = 3) matching and Mahalanobis distance matching. The absence rate in the treated schools is higher initially, but the differences are not significant.

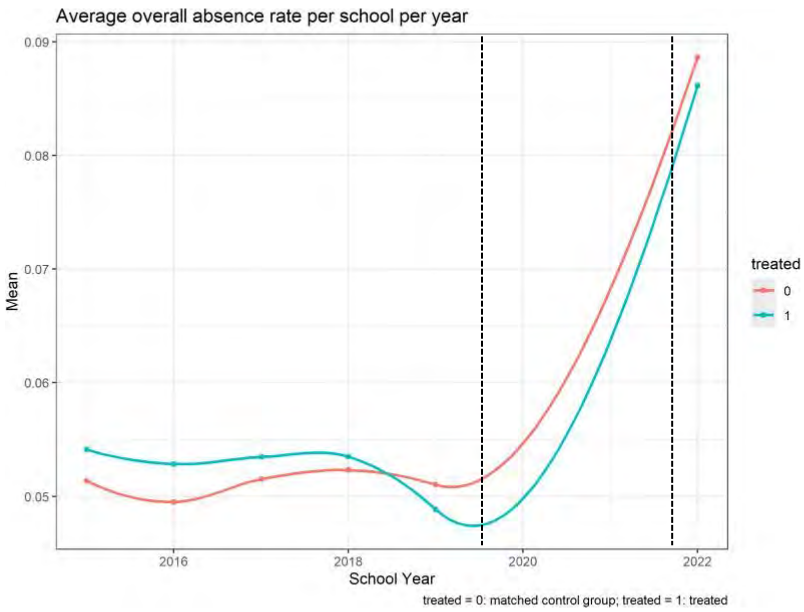
As the number of schools in these studies are limited, Tables 7, 8 and 9 show DiD regression estimates for absence rates, with the coefficient of interest being *Treatment × Post*. Estimates using linear models are included for completeness only and should be interpreted with caution, since the parallel trends assumption might be violated. The OLS model does not show any significant effect for the treated schools. Under a Poisson regression, neither estimate of interest is statistically significant at the 5 per cent level, and therefore, we are unable to conclude that absences decreased in participating schools at all.

Figure 7: Absence rate pre- and post-treatment – one-to-one matching



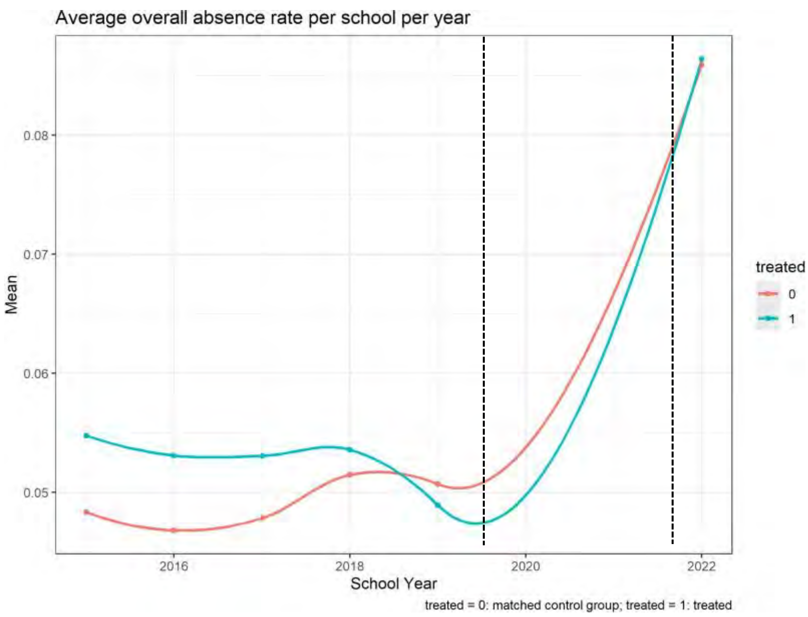
Notes: This figure shows the average absence rate in schools during the autumn and spring terms of each academic year in treated vs matched control schools using one-to-one matching (Source of data: ONS)..

Figure 8: Absence rate pre- and post-treatment – k-nearest neighbour (k = 3) matching



Notes: This figure shows the average absence rate in schools during the autumn and spring terms of each academic year in treated vs matched control schools using k-nearest neighbour (k = 3) matching (Source of data: ONS)..

Figure 9: Absence rate pre- and post-treatment – Mahalanobis distance matching



Notes: This figure shows the average absence rate in schools during the autumn and spring terms of each academic year in treated vs matched control schools using Mahalanobis distance matching (Source of data: ONS)..

Table 7: Difference-in-differences regression output for absence rates

	Absence rate	
	Linear	Poisson
Treatment × Post	0.001 (0.003)	0.009 (1.586)
School FE	Yes	Yes
Year FE	Yes	Yes
Sample size	164	164

Notes: The outcome variables are the overall absence rates (the absence rate in the autumn and spring term of each academic year) in the schools. The treated schools are participating schools, and the control schools are selected by one-to-one matching. * p < 0.10; ** p < 0.05; *** p < 0.01 (Source of data: ONS)..

Table 8: Difference-in-differences regression output for absence rates

	Absence rate	
	Linear	Poisson
On-route × Post	–0.003 (0.003)	–0.066 (1.473)
School FE	Yes	Yes
Year FE	Yes	Yes
Sample size	260	260

Notes: The outcome variables are the overall absence rates (the absence rate in the autumn and spring term of each academic year) in the schools. The treated schools are participating schools, and the control schools are selected by k-nearest neighbour (k = 3) matching. * p < 0.10; ** p < 0.05; *** p < 0.01) (Source of data: ONS)..

Table 9: Difference-in-differences regression output for absence rates

	Absence rate	
	Linear	Poisson
On-route × Post	–0.003 (0.003)	–0.066 (1.473)
School FE	Yes	Yes
Year FE	Yes	Yes
Sample size	183	183

Notes: The outcome variables are the overall absence rates (the absence rate in the autumn and spring term of each academic year) in the schools. The treated schools are participating schools, and the control schools are selected by Mahalanobis distance matching. * p < 0.10; ** p < 0.05; *** p < 0.01.(Source of data: ONS).

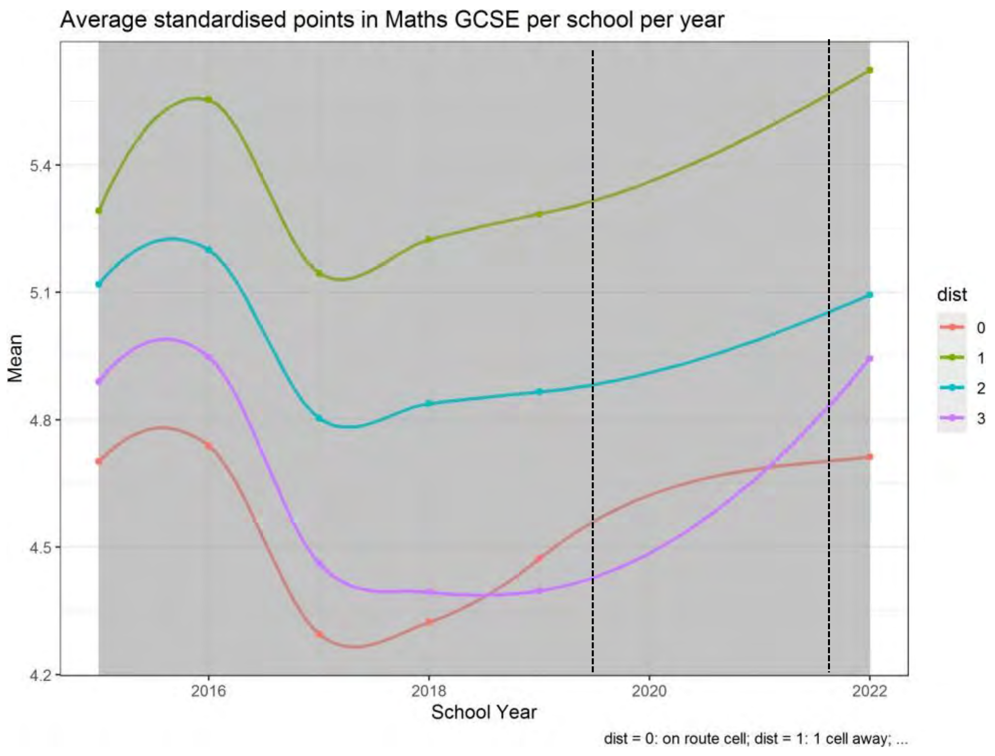
Student attainment – average standardised points in GCSE Maths

For this section, we have two outcomes. First, we study the average highest standardised points in GCSE Maths attained by students in the target schools. The highest standardised points in GCSE Maths creates an outcome variable between 0 and 10. The second outcome we are interested in is the percentage of students attempting GCSE Maths. This is shown in Figure A.3 in the Appendix.

Figure 10 shows the average standardised points in GCSE Maths in schools in each type of cell (i.e. in cells containing a treated route or one, two or three cells away) per academic year. The average standardised points in GCSE Maths in the treatment schools is lower in pre-treatment years. The figure shows that the average standardised points in GCSE Maths evolve in a parallel way both in the treatment schools (the schools located in cells where *dist* = 0) and the control schools (schools located in cells where *dist* = 1 and *dist* = 2).

Table 10 shows the DiD regression estimates for average standardised points in GCSE Maths, with the coefficient of interest being *On-route × Post*. Estimates using linear models are included for completeness only and should be interpreted with caution, since the parallel trends assumption might be violated. The OLS model does not show any significant effect for the schools located in treated cells. Given the abundance of zeros in our dependent variable (the average standardised points) we follow the statistics literature (Xiang, 2009; UCLA, 2024) and also develop a Poisson model to deal with the excess of zeros in our data, even though the average standardised points in GCSE Maths do not necessarily form a count variable. Under the Poisson specification, neither estimate of interest is statistically significant at the 5 per cent level, and therefore, we are unable to conclude that GCSE Maths scores increased in the schools along the routes.

Figure 10: Average standardised points in GCSE Maths pre- and post-treatment



Notes: This figure shows the average standardised points in GCSE Maths in an academic year for schools in each type of cell. Here, *dist* refers to how far a cell that a school is located in is from the nearest Step Together route. Schools where *dist* = 0 are located in cells on routes, while those where *dist* = 1 are located one cell away from a route (or equivalently, on an on-route cell) (Source of data:

ONS).

Table 10: Difference-in-differences regression output for average standardised points in GCSE Maths

	Standardised points in GCSE Maths	
	Linear	**Poisson**
On-route × Post	0.078 (0.081)	0.017 (1.007)
1-away × Post	0.212 (0.137)	0.035 (0.214)
2-away × Post	0.018 (0.104)	0.002 (0.170)
School Fixed Effect	Yes	Yes
Year FE	Yes	Yes
Sample size	1,260	1,260

Notes: Outcome variables are average standardised points in GCSE Maths in schools located in each type of cell. *On-route* refers to cells through which a Step Together route runs. *1-away and 2-away* refer to cells one and two cells away, respectively, from the nearest on-route cell. All standard errors are clustered by the Step Together route. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. (Source of data: ONS).

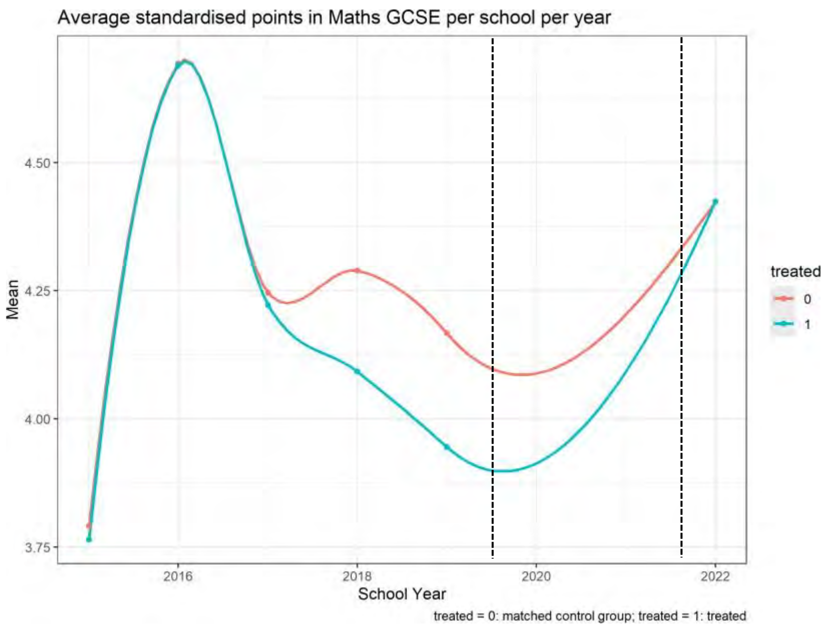
In the next step, we provide the result of the regressions, using three different matching techniques to select the control groups.

Figures 11, 12 and 13 show the average standardised points in GCSE Maths in treatment and control schools in each academic year. The treatment schools are the participating schools, and the control schools are selected using one-to-one, k-nearest neighbour matching (k = 3) and Mahalanobis distance matching.

Tables 11, 12 and 13 show DiD regression estimates for the average standardised points, with the coefficient of interest being *Treatment × Post*. Estimates using linear models are included for completeness only and should be interpreted with caution, since the parallel trends assumption might be violated. The OLS model does not show any significant effect for the treated schools. Under the Poisson regression, neither estimate of interest is statistically significant at the 5 per cent level, and therefore, we are unable to conclude that the average standardised points in GCSE Maths in participating schools increased at all.

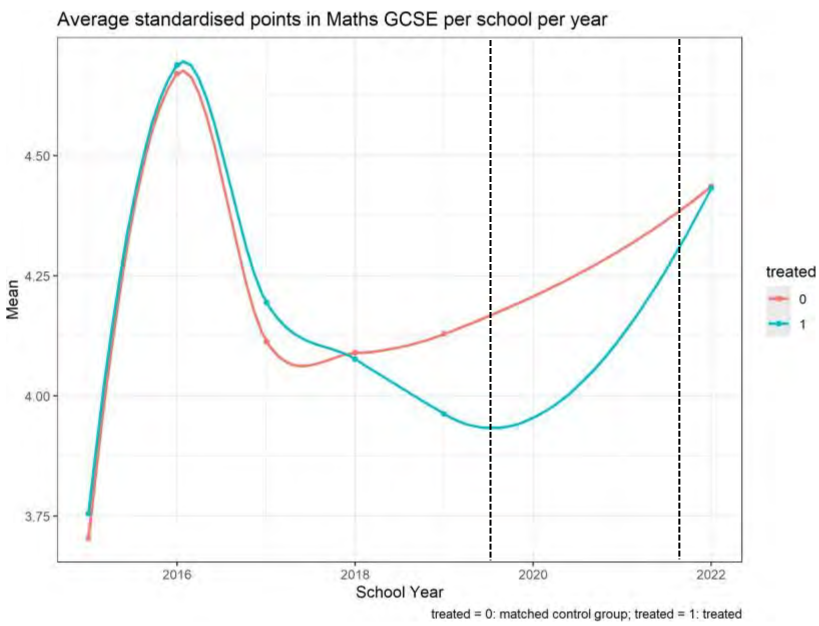
We also analyse the effect of treatment on the percentage of students who attempted GCSE Maths. This outcome shows diverging pre-treatment trends and violates the assumption of our research design. The results are shown in Figures A.3, A.4 and A.5 of the appendix. The analysis does not show any significant effect on the percentage of students attempting GCSE Maths.

Figure 11: Average standardised points in GCSE Maths pre- and post-treatment

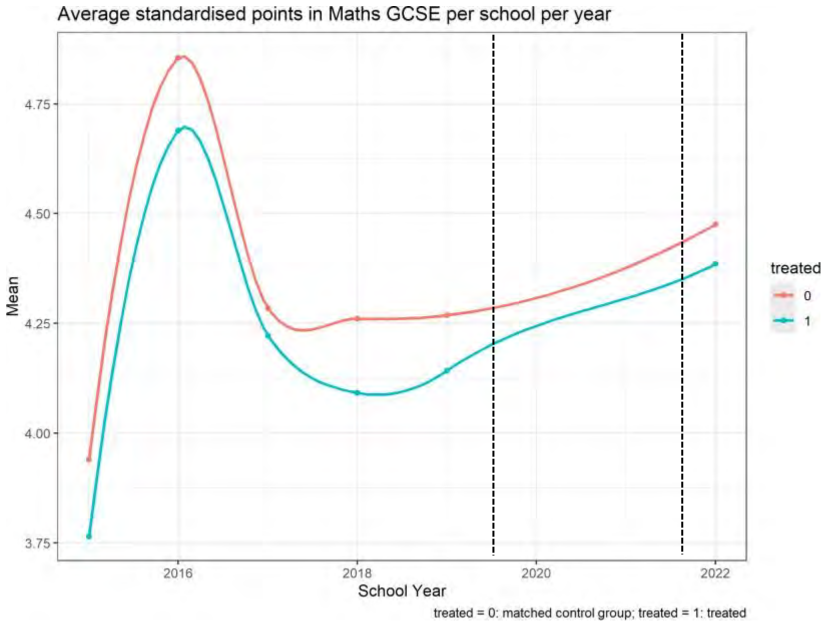


Notes: This figure shows the standardised points in GCSE Maths in each academic year and in treated vs matched control schools, using one-to-one matching (Source of data: ONS).

Figure 12: Average standardised points in GCSE Maths pre- and post-treatment



Notes: **This figure shows the standardised points in GCSE Maths in each academic year and in treated vs matched control schools, using k-nearest neighbour (k = 3) matching (ONSFigure 13: Average standardised points in GCSE Maths pre- and post-treatment (Source of data: ONS).**



Notes: This figure shows the standardised points in GCSE Maths in each academic year and in treated vs matched control schools, using Mahalanobis distance matching. (Source of data: ONS).

Table 11: Difference-in-differences regression output for average standardised points in Maths

	Standardised points in GCSE Maths	
	Linear	Poisson
Treatment × Post	0.116 (0.150)	0.027 (0.192)
School FE	Yes	Yes
Year FE	Yes	Yes
Sample size	183	183

Notes: The outcome variable is average standardised points in GCSE Maths in each school and academic year. The treated schools are participating schools, and the control schools are selected using one-to-one matching. * p < 0.10; ** p < 0.05; *** p < 0.001) (Source of data: ONS).

Table 12: Difference-in-differences regression output for average standardised points in GCSE Maths

	Standardised points in GCSE Maths	
	Linear	Poisson
On-route × Post	0.061 (0.102)	0.014 (0.014)
School FE	Yes	Yes
Year FE	Yes	Yes
Sample size	300	300

Notes: The outcome variable is average standardised points in GCSE Maths in each school and academic year. The treated schools are participating schools, and the control schools are selected using k-nearest neighbour (k = 3) matching. * p < 0.10; ** p < 0.05; *** p < 0.01 (Source of data: ONS).

Table 13: Difference-in-differences regression output for average standardised points in GCSE Maths

	Standardised points in GCSE Maths	
	Linear	Poisson
On-route × Post	–0.008 (0.105)	–0.001 (0.175)
School FE	Yes	Yes
Year FE	Yes	Yes
Sample size	204	204

Notes: The outcome variable is average standardised points in GCSE Maths in each school and academic year. The treated schools are participating schools, and the control schools are selected using Mahalanobis distance matching. * p < 0.10; ** p < 0.05; *** p < 0.01 (Source of data: ONS).

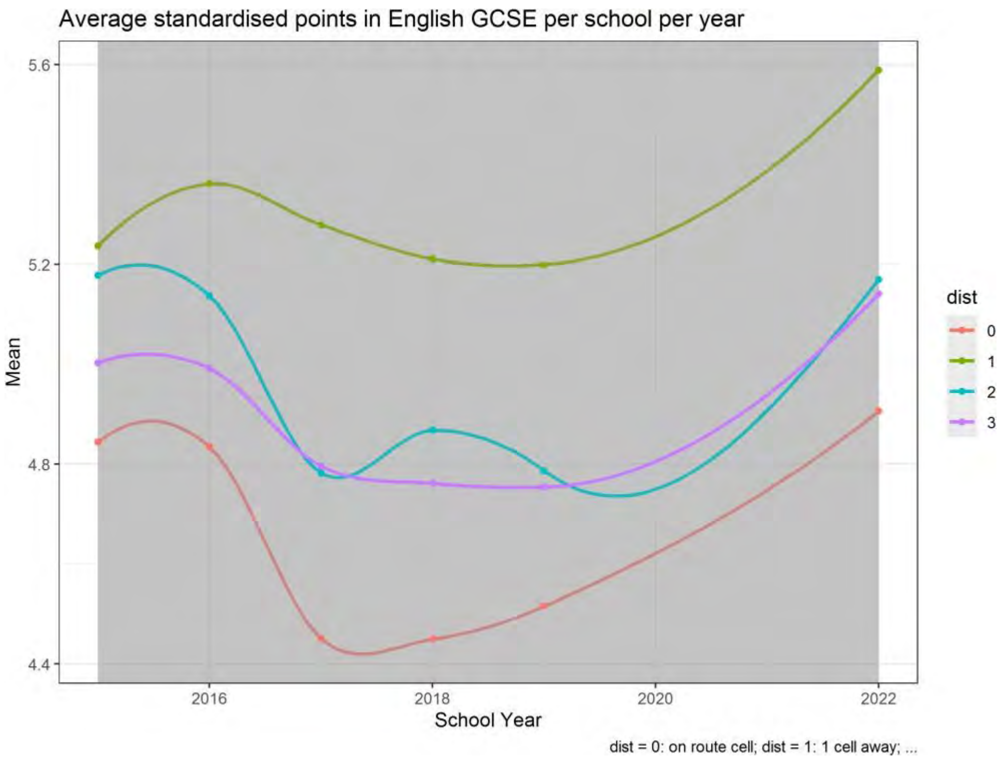
Student attainment – average standardised points in GCSE English

Similar to the previous section, in this section, we have two outcomes. First, we study the average highest standardised points in GCSE English attained by students in the target schools. The highest standardised points in GCSE English create an outcome variable between 1 and 9. Figure 14 shows the average standardised points in English in schools in each type of cell (i.e. in cells containing a treated route or one, two or three cells away) per academic year. Figure 16 shows the average standardised points in GCSE English pre- and post-treatment. The average standardised points in GCSE English in treatment schools is lower in pre-treatment years. Figure 16 shows that the average points in GCSE English evolve in a parallel way both in the treatment schools (the schools located in cells where *dist* = 0) and the control schools (schools located in cells where *dist* = 1 and *dist* = 2). The second outcome we are interested in is the percentage of students that attempted GCSE English. This outcome shows diverging pre-tends that violate the assumption of our research design. This is shown in Figure A.7 in the appendix.

Table 14 shows DiD regression estimates for average standardised points in English, with the coefficient of interest being *On-route × Post*. Estimates using linear models are included for completeness only and should be interpreted with caution, since the parallel trends assumption might be violated. The OLS model does not show any significant effect for the schools located in the treated cells.

We perform another exercise and use the schools located in cells more than three cells away from the treated cells as the control (see Table A.2). The pre-trends of GCSE English score attainment are visually parallel under this setting. We have again found a positive significant effect on GCSE English scores in the treated schools post-treatment and in this setting. However, given the challenges with the data, the the lack of consistent evidence for parallel trends, a lack of evidence on absence rates and/or feelings of safety, we cannot conclude that the treatment has had a positive overall effect on students' GCSE English scores in the affected schools.

Figure 14: Average standardised points in GCSE English pre- and post-treatment



Notes: This figure shows the average standardised points in GCSE English in an academic year for schools in each type of cell. Here, *dist* refers to how far a cell that a school is located in is from the nearest Step Together route. Schools where *dist* = 0 are located in cells on routes, while those where *dist* = 1 are located one cell away from a route (or equivalently, on an on-route cell) (Source of data: ONS).

Table 14: Difference-in-differences regression output for average standardised points in GCSE English

	Standardised points in GCSE English
	Linear
On-route × Post	0.193* (0.115)
1-away × Post	0.245 (0.195)
2-away × Post	0.142 (0.148)
School FE	Yes
Year FE	Yes
Sample size	1,260

Notes: Outcome variables are average standardised points in GCSE English Maths in schools located in each type of cell. *On-route* refers to cells through which a Step Together route runs. *1-away and 2-away* refer to cells one and two cells away, respectively, from the nearest on-route cell. All standard errors are clustered by the Step Together route. * p < 0.10; ** p < 0.05; *** p < 0.01 (Source of data: ONS).

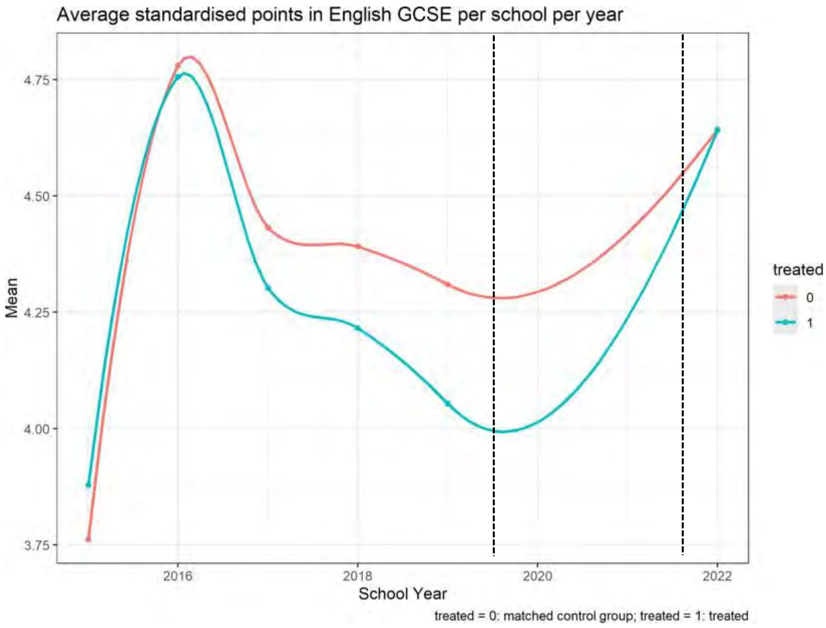
In the next step, we provide the results of the regressions, using three different matching techniques to select the control groups.

Figures 15, 16 and 17 show the average standardised points in GCSE English in the treatment and control schools in each academic year. The treatment schools are the participating schools, and the control schools are selected using one-to-one, k-nearest neighbour matching (k = 3) and Mahalanobis distance matching.

Table 15, 16 and 17 show DiD regression estimates for the average standardised points, with the coefficient of interest being *Treatment × Post*. Estimates using linear models are included for completeness only and should be interpreted with caution, since the parallel trends assumption might be violated. The OLS model does not show any significant effect for the treated schools.

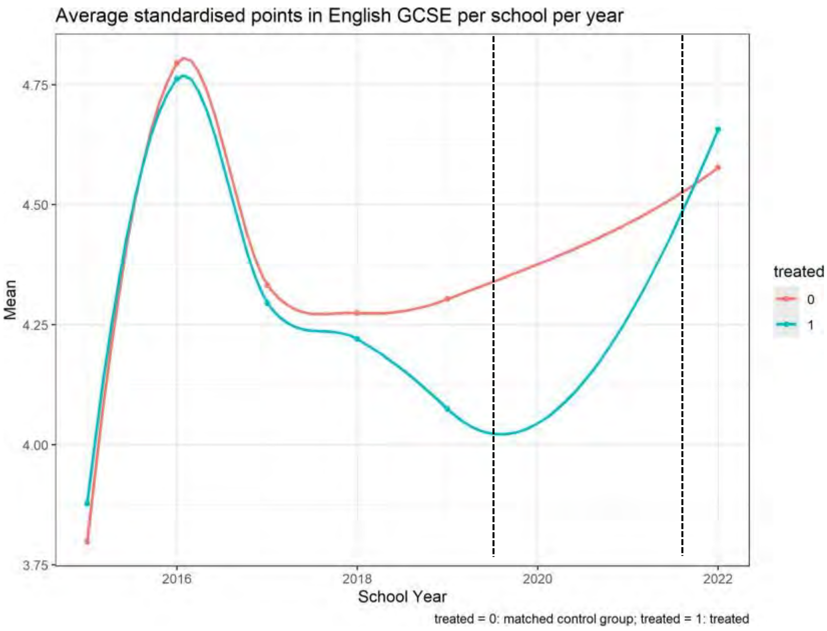
We also analyse the treatment effect on the percentage of students who attempted GCSE English. This outcome shows diverging pre-trends and violates the assumption of our research design. The results are shown in Figures A.6, A.7 and A.8 in the Appendix. The analysis does not show any significant effect on the percentage of students attempting GCSE English.

Figure 15: Average standardised points in GCSE English pre- and post-treatment



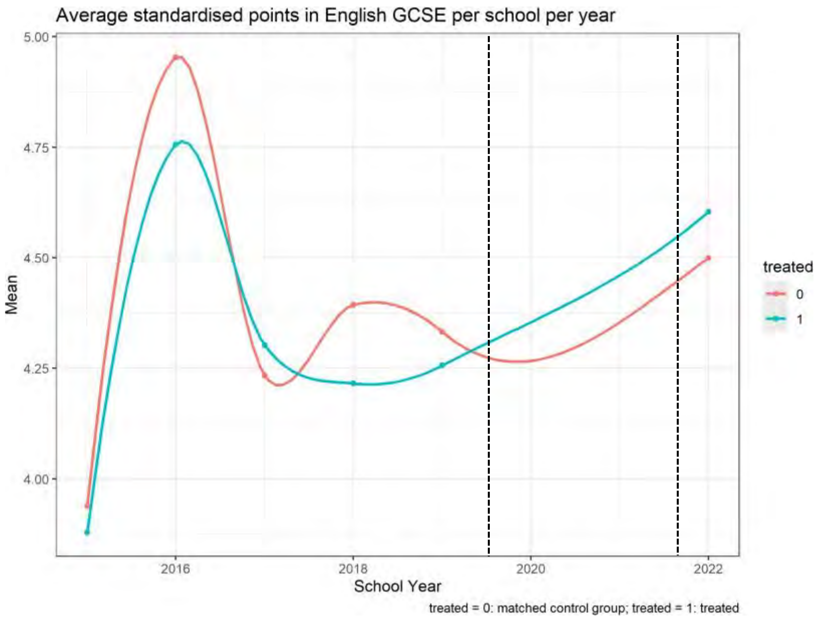
Notes: This figure shows the standardised points in GCSE English in each academic year and in treated vs matched control schools, using one-to-one matching (Source of data: ONS).

Figure 16: Average standardised points in English pre- and post-treatment



Notes: This figure shows the standardised points in GCSE English in each academic year and in treated vs matched control schools, using k-nearest neighbour (k=3) matching (Source of data: ONS).

Figure 17: Average standardised points in English pre- and post-treatment



Notes: This figure shows the standardised points in GCSE English in each academic year and in treated vs matched control schools, using Mahalanobis distance matching (Source of data: ONS).

Table 15: Difference-in-differences regression output for average standardised points in GCSE English

	Absence rate
	Linear
Treatment × Post	0.106 (0.155)
School FE	Yes
Year FE	Yes
Sample size	183

Notes: The outcome variable is average standardised points in GCSE English in each school and academic year. The treated schools are participating schools, and the control schools are selected using one-to-one matching. * p < 0.10; ** p < 0.05; *** p < 0.01 (Source of data: ONS).

Table 16: Difference-in-differences regression output for average standardised points in GCSE English

	Absence rate
	Linear
On-route × Post	0.181* (0.112)
School FE	Yes
Year FE	Yes
Sample size	300

Notes: The outcome variable is average standardised points in GCSE English in each school and academic year. The treated schools are participating schools, and the control schools are selected using k-nearest neighbour (k = 3) matching. * p < 0.10; ** p < 0.05; *** p < 0.01 (Source of data: ONS).

Table 17: Difference-in-differences regression output for average standardised points in GCSE English

	Absence rate
	Linear
On-route × Post	0.116 (0.185)
School FE	Yes
Year FE	Yes
Sample size	204

Notes: The outcome variable is average standardised points in GCSE English in each school and academic year. The treated schools are participating schools, and the control schools are selected using Mahalanobis distance matching. * p < 0.10; ** p < 0.05; *** p < 0.01 (Source of data: ONS).

Summary

Overall, we are unable to conclude that Step Together has had an impact on school absences or student attainment. The lack of precision in the estimates makes it difficult to say whether Step Together meaningfully affected educational outcomes, and while we do find a meaningful positive effect, the conditions required for the DiD method to be valid may not be met, and hence, the results may be biased.

Readiness for Trial

This study has conducted the evaluation of the pilot stage of the Step Together programme. We use a quasi-experimental design for the evaluation purpose. The findings do not reveal a meaningful impact by the Step Together programme in regards to reducing crime incidences or improving student attainment. While the study does not highlight any meaningful differences, it does provide evidence of promise, conditional on a careful consideration of the research limitations and of the risks of bias in future evaluations.

Step Together cost evaluation

Purpose

The LSE team is conducting a cost evaluation as part of the impact evaluation to measure the costs of delivering Step Together. This will help inform future decisions about the expansion and roll-out of the programme in other locations. For example, if another WMVRP were to try to determine whether the programme made sense for them, they could use this cost evaluation to understand exactly how much the programme would cost them and what it would cost in their local area under different circumstances (for example, if the size were different from that in the West Midlands). The programme was co-funded by the YEF and the Home Office.

Method

For the cost evaluation, we use the *bottom-up principal*: identifying the individual resources required to deliver an intervention, estimating the quantity of these resources needed and attaching monetary values to these resources. These figures are then combined to estimate the total amount spent on an intervention. We will estimate costs from the perspective of both the WMVRP and the providers.

Cost evaluation through the West Midlands Violence Reduction Partnership's lens

This section describes the cost for the WMVRP in the academic year 2021–2022 for the Step Together Evaluation project. Following YEF's cost guidance document, we identify the disaggregated costs and summarise them in a tabular format.

Table 18 describes the total costs incurred by WMVRP in the pilot phase of the Step Together project. The project continued to run after the pilot phase; however, this section highlights the costs for the pilot phase only. The cost items contain two types of costs: a) set-up costs (i.e. the up-front amount spent at the beginning of the project); and b) the recurring costs (i.e. the amounts that were paid repeatedly). The total staff cost incurred in the academic year 2021–2022 for the eight WMVRP staff was £72,345. This figure is made up of their salaries (i.e. £55,001) and the non-wage labour costs (including pension contributions and National Insurance), which was £17,344.. The salary and non-wage contributions were paid in two rounds: September–March and April–June.

Table 18: West Midlands Violence Reduction Partnership (WMVRP) costs, academic year 2021–2022

Cost items	Upfront/Set up or recurring	WMVRP £	Chaperone £	Total £
Staff				
WMVRP	Total wage disbursed	72,345		72,345
Chaperone	Set-up		92,529	92,529
Chaperone	Recurring		946,287	946,287
Other costs				
Additional £3K per route allocated to 18 routes	Set-up	54,000		54,000
Evaluation	Recurring	174,977		174,977
Total costs				
Set-up				1,46,529
Recurring				1,193,609
Total costs				1,340,138

A total of eight members from WMVRP were involved in the Step Together project for its 39-week duration. On average, WMVRP staff dedicated 13.4 hours per week to the Step Together project.[16] The staff duties throughout the year, broken down by staff members, have been summarised in Table 19 as they relate to the Step Together delivery. Depending on the type of task carried out by a WMVRP staff member, the per day labour cost ranged from £130 to a maximum of £650.

The WMVRP released funds to 10 providers in a cycle that managed 18 routes as part of the Step Together programme.[17] Each provider or site deployed chaperones on their routes and had additional staff members for the purpose of management. The WMVRP disbursed funds to each provider at the beginning of the three terms (i.e. autumn, spring and summer), which are attributed as recurring costs. This amount was given to meet the financial requirements at each site in terms of wages for staff and chaperones, project management and any other recurring expenses.

For the academic year 2021–2022, WMVRP provided the 10 sites with £580,181 for the first two terms and £366,106 for the last term. As indicated in Table 18, this amount summed up to £946,287. The chaperone set-up cost of £92,529 was incurred by WMVRP at the beginning of the academic year to allow sites/providers to put devices, uniforms and phones in place, along with meeting any other material or equipment expenses. Along with the chaperone costs, an additional amount of £3,000 was allocated to each of the 10 sites and 18 routes at the beginning of the academic year to help with the set-up expenses. For this purpose, the WMVRP incurred a total cost of £54,000, as highlighted in Table 18. A total of £174,977 was given in kind by the YEF for the purpose of the Step Together project evaluation. The total set-up or upfront cost for WMVRP was £146,529, and the recurring costs amounted to £1,193,609. Thus, the total cost for running one cycle of the Step Together pilot project using the outlined model of delivery was £1,340,138.

[16] Average = total number of hours / total number of WMVRP staff

[17] An additional (nineteenth) route was launched during the life of the pilot. It has not been included in this evaluation.

Table 19: Staff duties – academic year 2021–2022

Role	Description of staff duties	Number of staff	Total hours per week
Education delivery manager	Supports the development and delivery of Step Together. Provides direct line management for the education navigator.	1	14.36
Education navigator	Responsible for the day-to-day operations of the Step Together programme. This includes liaising with the schools and providers to ensure its successful delivery.	1	28.72
Governance assistant	Provides administration support to all aspects of the programme.	1	7.18
Data analyst	Responsible for analysing data to ensure the routes selected are appropriate and provide the opportunity to reduce the most risk to young people.	1	7.18
Strategic support	WMVRP SLT strategic support	2	3.59
Navigator	Navigator	1	7.18
Commissioning officer	Prepares the grant applications and works with the business support team to resolve purchase order issues and ensure the payment of invoices.	1	35.9
Total staff		8	
Average number of hours spent			13.4

Finally, we analyse and present the cost per participant for one cycle in Table 20. The cost per participant has been obtained by dividing the total cost by the total number of participants. We also measure the set-up cost per participant and the recurring cost per participant following the same method. The total number of participants in the Step Together project across the 18 routes where it operated was 18,000 for the academic year 2021–2022. The set-up cost per delivery cycle, per participant was £8 and the recurring cost per delivery cycle, per participant was £66. The total cost per participant was £74 (= 1,340,13873 / 18,000). It is worth noting that this amount has been identified as the cost for this particular model of delivery, which had 18 routes with up to four and as many as 10 chaperons managing a route.

Table 20: Cost per participant – academic year 2021–2022

No of participants	18,000
Set-up costs per participant	8
Recurring costs per participant	66
Total costs per participant	74

Cost evaluation through the provider's lens

This section describes the cost to run the Step Together programme incurred by the sites/providers in the academic year 2021–2022.[18] After receiving the funds from WMVRP, the sites/providers exercised discretion regarding their fund allocation decisions. Following a similar evaluation strategy as above, we decipher how the funds were allocated to various purposes to run the programme. Detailed information on the costs for the cycle was obtained from a survey of the sites. Table 21 provides the details of each aggregated cost for the sites and for the academic year 2021–2022. To ensure the privacy of staff information, the surveys were conducted anonymously. There were 10 providers/sites managing 18 routes, and we were able to gather responses from nine of them. We use extrapolation based on mean values to provide an estimate for the 10th provider (we take similar approach in case of missing information).

Table 21: Site/provider costs – academic year 2021–2022

Cost Items	Upfront or recurring	Chaperone £	Total £
Staff			
Total staff costs (managers, workers, chaperones)	Recurring		780,332
Chaperone			
Non-wage labour costs	Set-up	11,577	28,943
Training	Upfront or recurring	16,200	16,200
		Programme	
Bespoke learning manuals			NA
		Buildings and facilities	
Venue/office rent	Recurring		3,150
Procurement costs	Upfront	2,122	2,122
		Materials and equipment	
Materials (uniforms, stationary, printing, badges)	Recurring		16,816
		Other costs	
WMVRP staff cost	Total wage disbursed		72,345
Additional £3K per route allocated to 18 routes	Set-Up	54,000	54,000
Evaluation	Recurring	174,977	174,977
		Total costs	
Set-up			85,064
Recurring			1,063,820
Total costs			**1,148,884**
Difference to WMVRP amount			**191,254**

[18] The terms site/provider are used interchangeably.

Table 22: Provider costs – academic year 2021–2022

Site no.	Total staff cost across the 18 routes £	Number of weeks	Total £
1	2,500	39	97,500.00
2	2,000	39	78,000.00
3	2,600	39	101,400.00
4	1,500	39	58,500.00
5	1,500	39	58,500.00
6	1,989.72	39	77,599.08
7	3,000	39	117,000.00
8	2,500	39	97,500.00
9	Not provided	39	16,300.00
10	Assuming the average of the total amount from the nine sites to be the total cost for the 10th site	N/A	78,033.23
Total staff costs for the 18 routes			780,332.31
Average staff costs			78,033.23

The total staff cost incurred by the sites to manage the routes in one cycle aggregated to £780,332. This included the cost for management, chaperones and other staff members. The breakdown for the 10 sites is summarised in Table 22. On average, the staff cost across the 10 sites came down to approximately £78,033 per cycle, or £2,000 per week. The weekly staff cost ranged from £1,500 to £3,000. The variation in costs can be attributed to differences in the number of routes managed by a provider.

To estimate the non-wage labour costs, we asked the sites/providers to say whether there were any non-wage labour costs in the survey. Among the nine respondents, five paid out non-wage labour costs during the delivery cycle. However, information about the allocated amount was only given by two providers. This amount summed to £11,577.50. The other three sites/providers chose not to disclose the amount in the survey. To deal with the missing information generated from the providers' incomplete response, we estimate an approximate amount for the three sites using the mean values ([£11,577.5 / 2] × 3 = £17,366.25). Thus, the total non-wage labour cost incurred during the 2021–2022 delivery cycle was £28,943.

We also collected information on training costs. Of the nine sites surveyed, six sites used WMVRP funds to train the chaperones. Two sites either provided free-of-charge training or outsourced it every month and thus did not use WMVRP funds. One site did not incur any training costs. Among the six sites that used WMVRP funds for training, five sites mentioned the training amount exclusively, while the last included it as part of the staff costs. We add the total cost incurred across five sites. The sixth site's cost have been intentionally left out to avoid double counting. Thus, a total of £16,200 was allocated for training purposes across five different sites in the 2021–2022 cycle. The building and facilities costs for this cycle were £3,150 and £2,122 for venues and procurement, respectively. Among the sites surveyed, three sites paid to rent an office or venue from the funds given by the WMVRP. The programme procurement costs were incurred by only two sites.

Finally, the providers used £16,816 to obtain materials and equipment in 2021–2022. In the survey, seven sites provided information on the materials and equipment costs (including uniforms, stationery, travel, printing, etc.), which totalled £14,714. Another site suggested that it had used funds to pay for uniform and travel costs; however, it failed to provide an exact estimate. We use the average value of the seven sites (i.e. £2,102) as an estimate for this value. Thus, we arrive at the total figure, which is £16,816 (£14,714 + £2,102). Finally, we add the WMVRP staff cost, the evaluation cost and the additional funds as other costs (as per Table 18). The total costs from the sites/providers' perspective were £1,148,884. In other words, on average, it costs £63,826 for a site/provider to operate a route. The total set-up or upfront cost for providers was £85,064 and the recurring costs were £1,063,820.

Conclusion

Table 23: Summary of feasibility study findings

Research question	Findings
RQ1. Can we collect the necessary crime and education data within the initial time frame set out for the evaluation?	**Crime data** It was possible to map crime data to routes; however, we were unable to obtain the necessary data within the initial time frame because of challenges in resolving data sharing agreements and discrepancies as a result of administrative system changes. **Education data** We were able to access the NPD with approval from the Department for Education. However, some schools were missing from the dataset for certain academic years. Since the missing data appeared to be random, we proceeded with the analysis. The number of observations for each study (attainment and school absences) varies and is determined based on the availability of the relevant variables. Additionally, attendance data is missing for all schools in the summer term of the 2021–2022 academic year. As a result, the total absence rate for that academic year will be unavailable. To address this, the total absence rate for the first two terms will be used to mitigate this data gap. Finally, we did not have access to academic year 2022–2023 as the ideal post-treatment academic year. It is advisable to incorporate the data from this year for a more precise future study.
RQ2. How can a plausible counterfactual be constructed?	**Counterfactual for crime** We used cells next to routes as a counterfactual. This resulted in a plausible counterfactual for violent and property crime, as shown by the lack of pre-trends in dynamic difference-in-differences (DiD) regression results. However, for antisocial behaviour (ASB), neighbouring cells are not a plausible counterfactual. Alternative approaches, perhaps including matching, should be considered. Our parallel pre-trends analysis excluded two years of crime data corresponding to the COVID-19 pandemic period from 2020–2021, when crime rates were unusual. Even if the neighbouring cells were a plausible counterfactual pre-pandemic (i.e. from 2016–2019), it is unclear whether they would remain plausible counterfactuals post-pandemic. **Counterfactual for education** We designated the participating schools as our treatment group and applied matching techniques to select control schools from non-participating schools in the West Midlands. Additionally, we conducted an extended analysis to examine a broader range of schools. In this exercise, schools located in cells overlapping with a route were classified as the treatment group, while those in adjacent cells were used as controls. The adjacent cells serve as the plausible counterfactuals for the analysis. This approach was taken due to the limited number of participating schools, the small geographic area where the intervention was applied and the resulting constraints in generating matched samples.

RQ3. Is there a meaningful difference in crime and/or ASB along chaperoned routes compared to that along controls?	We cannot conclude that the chaperone presence resulted in a meaningful difference to violent and property crime, although this is likely to be due to low statistical power. We were unable to establish whether there was an effect of Step Together on ASB, given the lack of evidence for parallel trends in ASB in the treatment and control areas.
RQ4. Is there any evidence of spatial and/or temporal spill over (displacement) of crime or harm to the areas adjoining routes?	There is no evidence of spillover for violent and property crime. However, the standard errors are fairly large due to a lack of statistical power, and the possibility that the parallel trends assumption may not hold means that these estimates should be interpreted cautiously.
RQ5. Do schools with chaperones show decreased student absences (and therefore higher attendance rates) and higher student attainment compared that seen in to control schools?	We were unable to establish whether there was an impact on absences (and, therefore, on attendance). We were also unable to determine whether GCSE attainment increased as a result of lower absences (increased attendance), given the challenges with the data and the lack of evidence for parallel trends.
RQ6. Do schools with chaperones show increased feelings of safety among pupils, teachers and parents?	At this stage, we are unable to answer this question due to the extremely low response rate (10%) for the surveys rolled out to examine this question.

Evaluator judgement of intervention and evaluation feasibility

The main stage evaluation can be particularly improved by mitigating the risks associated with a lack of data.

- There were significant delays in receiving access to the WMP crime data, which was needed to measure the primary outcome. Inconsistencies in the original data detected during the initial analysis led to recollection of the data from WMP. This led to considerable delay in the process of evaluation. To mitigate the risks around this issue, it is suggested that the involved parties co-design the information-sharing agreement. Moreover, an estimated overview of the timeline for data availability should be provided, while early meetings with the evaluation team would make the process more efficient for future evaluations.

- Measuring the secondary outcome (student attainment) was conditional on LSE getting access to the NPD data. Data from the DfE was released much later than expected, and the lack of multiple analytical packages on the ONS's server (where the data was stored) caused significant backlog affecting the completion of the analysis for the secondary outcome. To mitigate risks around this issue, future evaluations should carefully consider and allocate elongated time periods for unforeseen delays.

Interpretation

Step Together is a pilot programme in the West Midlands that aims to provide students with access to a trusted adult as they travel to and from school. The main objective of the pilot was to reduce crime and ASB and decrease school absences in areas with a chaperone presence. This intervention is similar to the evidence-based Safe Passage scheme in Chicago, Illinois. Step Together was delivered across approximately 18 routes in the West Midlands. These routes, with an average of six chaperones per route, monitor high footfall routes for at least two hours per day, at peak times before and after school. There are several lessons that could be learned from the impact evaluation and considered for future purposes.

- The important challenges this evaluation faced in analysing the effect of Step Together on crime and ASB were the risk of bias estimation and the lack of statistical power. The former issue is highlighted through the lack of a proper counterfactual for estimating the effects, as shown through the violation of the parallel trends assumption. The second issue is demonstrated by the large estimated minimum detectable effects (see the appendix). In spite of the large minimum detectable effect, point estimates for ASB and property crime are fairly large. The analysis reveals a decline for ASB and property crime of around 19 and 8 per cent, respectively. At the same time, there is no evidence for displacement of crime and ASB away from routes, with point estimates for spillover effects being either negative or positive, but nonetheless small. Overall, although most of our estimates are not meaningfully different, there is some promising evidence.

- However, the two most significant challenges in analysing the effect of Step Together on crime and ASB come from the lack of evidence for the parallel trends assumption and the potential effects of the pandemic. Additional but still significant challenges include missing data on compliance/fidelity and on pupil feelings of safety (which is the key causal mechanism of improved attendance).

- The findings from this study contrast with Chicago's Safe Passage programme. The estimates from the Safe Passage programme indicate smaller point estimates for property crime than violent crime, which is the opposite to those in the current study. Moreover, the standard errors were smaller in the analysis of the Safe Passage programme. However, it is important to consider that youth crimes in the US are different from those in the UK, due to a number of institutional, social and economic factors. Thus, while the two programmes seem comparable, the context for each differs widely. This should be kept in mind in future evaluations.

- The second part of the study is focused on student absences and attainment. To sum up, the study does not show that the Step Together programme had any significant effect on the student absence rate. We find a negative effect on absence rates in schools in adjacent areas, which suggests a spatial spillover of the programme. There is no significant effect on GCSE Maths and English attainment in schools in the treated and adjacent areas.

The extent to which the pilot supports the logic model

The pilot study follows the logic model as outlined by the Ipsos implementation and process evaluation team. As highlighted in the logic model/theory of change, Step Together was launched based on a series of hypotheses. The activities from this intervention were expected to lead to certain outcomes for crime, feelings of safety and education. This report analyses a series of hypotheses using six research questions highlighted in the Research Questions section. The logic model hypothesised that the Step Together programme would lead to:

- **A reduction in crime and ASB in the identified routes**. This has been answered by RQ3 ('Is there a meaningful difference in crime and/or ASB along chaperoned routes compared to that along controls?') Our study cannot conclude that chaperone presence results in a meaningful difference to violent and property crime, although this is likely due to low statistical power. The analysis suggests that chaperone presence does result in a meaningful reduction in ASB.

- **Potential reduction in crime and ASB in adjoining routes.** This has been answered by RQ4 ('Is there any evidence of spatial and/or temporal spillover (displacement) of crime or harm to the areas adjoining routes?') Our study cannot conclude whether there was a meaningful impact on ASB or not, given a lack of evidence of parallel trends. However, there is some evidence of reduced ASB in the areas near routes, suggesting positive spillover effects.

- **A decrease in school absences (and therefore an increase in school attendance) and an increase in student attainment in participating schools.** This has been answered by RQ5 ('Do schools with chaperones show decreased student absences (and therefore higher attendance rates) and higher student attainment compared that seen in to control schools?') Our study cannot conclude whether there was a meaningful impact on absences or not, given a lack of evidence of parallel trends. We also could not verify whether there had been a positive effect on students' GCSE scores in affected schools.

- **Increased feelings of safety.** This was covered by RQ6 ('Do schools with chaperones show increased feelings of safety among pupils, teachers and parents?'). At this stage, we are unable to answer this question due to an extremely low response rate (10 per cent) to the surveys rolled out to examine this question.

The generalisability (applicability) of the pilot study methods and findings to the future efficacy study and other potential studies.

The key method used in this pilot study (namely, using nearby cells as controls) was adapted from existing peer-reviewed economics studies. The methods used in the pilot study are therefore widely applicable to future studies. The findings, however, are not as widely generalisable.

- Large standard errors in some cases (particularly for property crime) mean that negative effects on crime may exist, even though there was insufficient evidence in this pilot study. Where there are potentially violations of the parallel trends assumption, those results should be interpreted with caution and should not be generalised to other potential studies.

- Another issue is that crime data only reflects *reported* crime, not actual crime. It is possible, for example, that Step Together patrols reduce actual crime but also increase reporting; hence, reported crime remains unchanged. This is a challenging issue to address, although it would likely be helpful if future evaluations are able to successfully survey students on feelings of safety and/or the crime they have experienced.

- The fact that there has been limited number of schools in the area that we conducted the study, in addition to the fact that we have not had access to a comprehensive dataset able to capture the diverse observable characteristics of the schools on a yearly basis) has limited the generalisability of the pilot study. The inclusion of other observable characteristics in an estimation exercise will allow us to capture the effect from the programme without the risk of any bias that other characteristics can create. Such biases confound the result and can lead to an over- or underestimation of the effect. Thus, it is important to include these biases in future estimation exercises.

Limitations of the pilot study

- The main limitation of this study is around the plausibility of counterfactuals. While neighbouring cells appear to be a plausible counterfactual for crime, for other outcome variables, the plausibility is somewhat questionable (demonstrated, for example, by an inspection of pre-trends for ASB). This situation may lead to bias in the regression results. Further data collection, perhaps with a treatment period longer than one academic year, could help to clarify the plausibility of counterfactuals.

- The analysis assumes that every route is treated for the entire 2021–2022 academic year.[19] In reality, the routes began at different times, with the last route (Kingstanding) launching in March 2022. Hence, estimates are likely to be biased towards zero, since parts of the 'treatment period' are actually pre-treatment. In addition, the treatment period outcomes might be contaminated by the COVID-19 pandemic; that is, the pandemic may have disproportionally affected the absence rates in the treated and control schools.

- Regarding education outcomes, the main limitation is also the identification and creation of valid counterfactuals using the limited data that we have access to. While we find a relationship between the presence of chaperones and students' attainments, this finding might be prone to bias due to a violation of the parallel trends assumption.

- Moreover, the GCSE marking scheme was affected by the pandemic in the post-treatment period; if this effect is different in the control and treatment schools, it will create an issue for our identification.

- Additionally, some routes only took place for part of the academic year, which is a limitation. This is particularly an issue for absences, as we probably have only one or two terms' data with which to track absence in affected schools.

- Future studies also need to carefully consider the time that might potentially be taken up by legal agreements, admin work and data inspection and sharing before setting up the timeline for analysis. Moreover, the involvement of different parties in data provision and in reviewing the analysis can substantially affect the timeline for the evaluators. Thus, a carefully constructed timeline that accounts for all these details is needed before an initial time frame is agreed. Ideally, a time period of two-and-a-half years is required for data collection, analysis, evaluation and reporting.

Future research and publications

Suggestions for future study methodology

The recommendations for future research include the following:

[19] The study plan was developed under the assumption that treatment would start simultaneously across all routes.

Evaluation design

- Future studies should consider incorporating randomisation to ensure appropriate counterfactuals are available. This is an important recommendation, as the current study is not able to incorporate an appropriate comparison group, which makes it difficult to interpret results without bias.
- If randomisation is not feasible, it is recommended that future evaluators use a more extensive and broader dataset that contains information on observable characteristics. These characteristics can be used as controls in a quasi-experimental design (such as DiD) to avoid the confounding of the results, especially when testing the impact on student attainment. In the current sample, the number of schools was limited to 24. A broader dataset also opens up an avenue for another method of identification by using matching techniques to find an appropriate counterfactual.
- In addition to the above, surveys can be done to study the general feeling of safety for teachers and staff and how they assess the effect of the programme on their students beyond the scope of school absences and GCSE results.
- To capture the effect on individual students, future studies need access to data with longer pre- and post-treatment periods, which would allow them to develop a valid counterfactual. This counterfactual could be constructed by focusing on 'movers' – students who were initially in an untreated school and later moved to a treated school, or vice versa. This approach is only feasible if researchers have the data to track students for multiple years following the intervention.
  - Additionally, a longer post-treatment period would allow researchers to better account for the long-term effects of the pandemic on educational outcomes, including students' school absence patterns. In an extended post-treatment period, outcomes are more likely to stabilise and return to pre-pandemic patterns, enabling a more valid and reliable evaluation of the intervention's impact.

Delivery and compliance/fidelity monitoring

Delivery of the programme: As shown by the implementation and the process evaluation by IPE and the pilot evaluation conducted by us, the Step Together programme was delivered as per the plan and the theory of change highlighted in IPE's report. The recommendations in our evaluation are the same as those of IPE, with nothing further to add.

The initial study plan did include an attempt to design and roll out a GPS application to track the route of chaperones. This would have allowed us to track the routes of the chaperones and measure compliance. For this purpose, LSE designed the app and commissioned YEF to launch it. However, the GPS app was more difficult to roll out (i.e. GPS data was limited) than anticipated. The main problems were threefold 1) the application was difficult to be installed on phones that were old and unsupportive of the app, 2) at most times chaperones did not turn on the application for use, and 3) finally, among those who did turn it on (a very small fraction), some didn't walk the prescribed route. These issues prevented us from capturing using the data from the application and thus measuring the compliance to treatment. We recommend attempting to collect this in future evaluations in order to understand if the chaperones stick to their assigned routes, which we assume is the case in our impact analysis, or if they regularly opt to make modifications to their routes, in which case we would also conduct our impact analysis using different treatment areas.

Outcome measures

The data collected for pre- and post-intervention on the basis of the appropriate delivery of the programme allowed us to evaluate the main outcomes: crime and educational attainment. However, low response rates to surveys prevented us from assessing the impact on feelings of safety. For future studies, it is recommended that evaluators print physical copies and have someone present in the schools to administer them.

Future research questions

The future research questions to be prioritised include:

- Do patrols reduce ASB? Is there any temporal displacement of crime/ASB (for example, from school time to night-time)? This research question will help to understand the effect of the Step Together programme on immediate crime outcomes (i.e. ASB), due to an increase in the presence of patrols in the area. Moreover, studying the temporal displacement will help to understand the shifts in crime patterns that may happen due to patrolling.

- Do patrols increase general feelings of safety for students, teachers and staff? How does this programme affect students beyond the scope of school absences and GCSE results? (Examples of other outcomes might be improvements to the students' motivation or general feelings about their school, or encouragement to engage in social activities or sports activities with school peers out of school hours – which would require them to commute to adjoining areas.) While this research question could not be studied during the pilot due to a lack of responses to the survey, it remains a key question to be studied to understand the mechanism behind the effect. It also documents the first-hand effect that an intervention like Step Together can create.

References

BBC News. (August 22, 2024). GCSE grades 2024: The 9–1 boundaries explained. https://www.bbc.co.uk/news/education-48993830

Bierman, R. et al. (2023, March). Step Together pilot: Implementation and process evaluation. Youth Endowment Fund. https://youthendowmentfund.org.uk/funding/evaluations/step-together/

Burlig, F., Preonas, L., & Woerman, M. (2020). Panel data and experimental design. *Journal of Development Economics, 144*, 102458. https://doi.org/10.1016/j.jdeveco.2020.102458

Gonzalez, R., & Komisarow, S. (2020). Community monitoring and crime: Evidence from Chicago's Safe Passage program. *Journal of Public Economics, 191*, 104250. https://doi.org/10.1016/j.jpubeco.2020.104250

Holm, S. (1979) "A simple sequentially rejective multiple test procedure".Scandinavian Journal of Statistics, 6(2), 65–70.

Home Office (2018, April). *Serious violence strategy*. Gov.UK. https://www.gov.uk/government/publications/serious-violence-strategy

Jadhav, C. (2018). GCSE 9 to 1 grades: a brief guide for parents. *The Ofqual Blog*. https://ofqual.blog.gov.uk/2018/03/02/gcse-9-to-1-grades-a-brief-guide-for-parents

McMillen, D., Sarmiento-Barbieri, I., & Singh, R. (2019). Do more eyes on the street reduce crime? Evidence from Chicago's safe passage programme. *Journal of Urban Economics, 110*, 1–25. https://doi.org/10.1016/j.jue.2019.01.001

Office for National Statistics (2022, April 28). Knife or sharp instrument offences. *Crime in England and Wales: Year ending December 2021*. https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingdecember2021

Department for Education, Department of Work and Pensions, Higher Education Statistics Authority, HM Revenue & Customs (DATE), National Pupil Database bespoke extract, Office for National Statistics, dataset

Romano, J. P., & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, *73*(4), 1237-1282.

UCLA. (2024). Poisson regression: R data analysis examples. UCLA Advanced Research Computing. https://stats.oarc.ucla.edu/r/dae/poisson-regression/

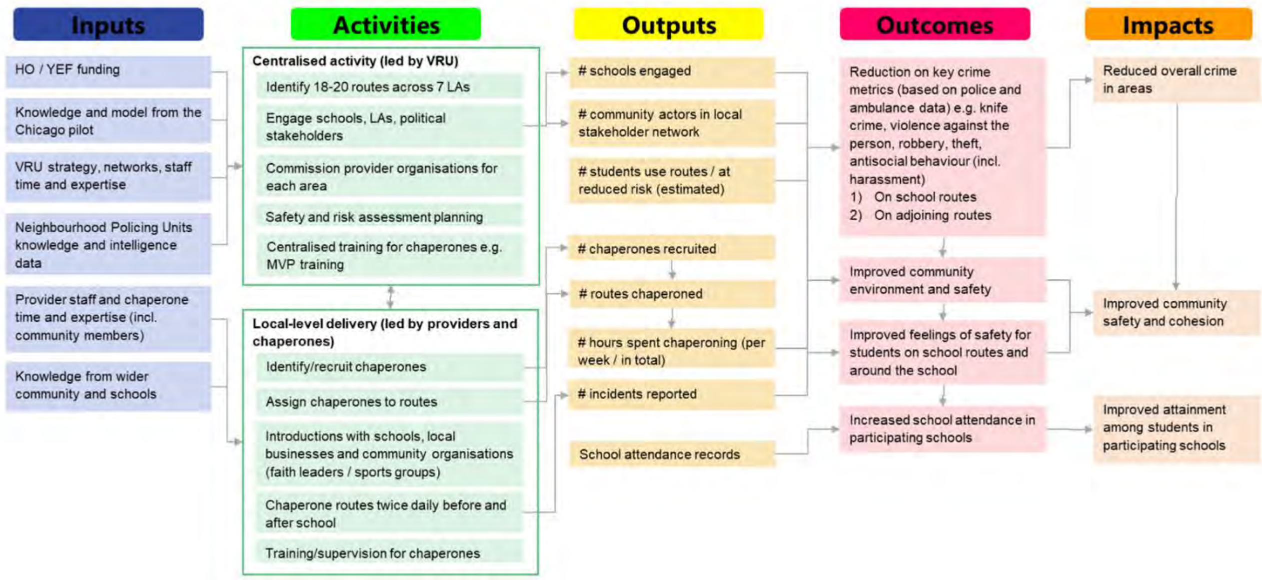
Vulliamy, P. et al. (2018). Temporal and geographic patterns of stab injuries in young people: A retrospective cohort study from a UK major trauma centre. *BMJ Open 8*, e023114. https://bmjopen.bmj.com/content/8/10/e023114

Xiang, A. (2009). An introduction to count data models. *Harvard Business School.*

https://www.hbs.edu/research-computing-services/Shared%20Documents/Training/count_data.pdf

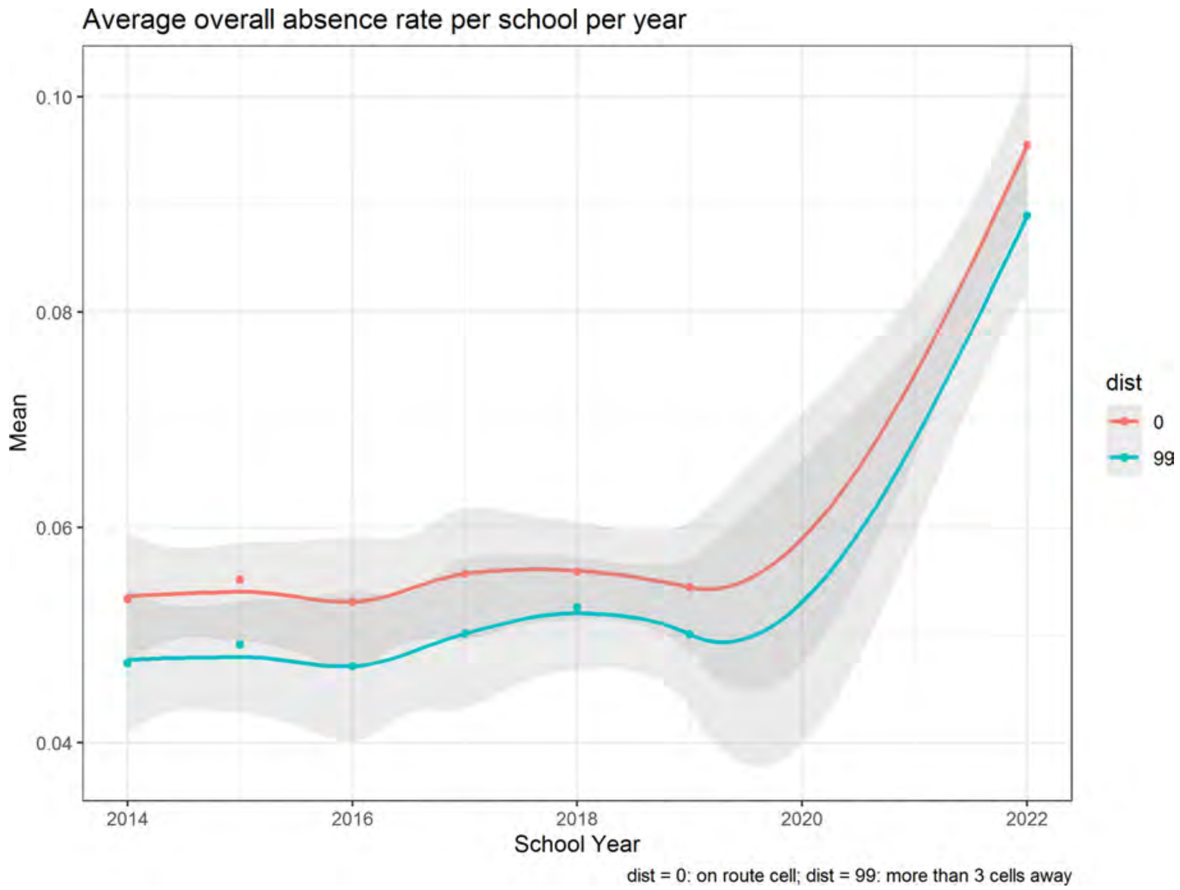
Appendix

Figure A.1: Step Together pilot logic model/theory of change



Notes: This figure shows the logic model developed by IPSOS for Step Together. (Source: Evaluation Report- Step Together Pilot)

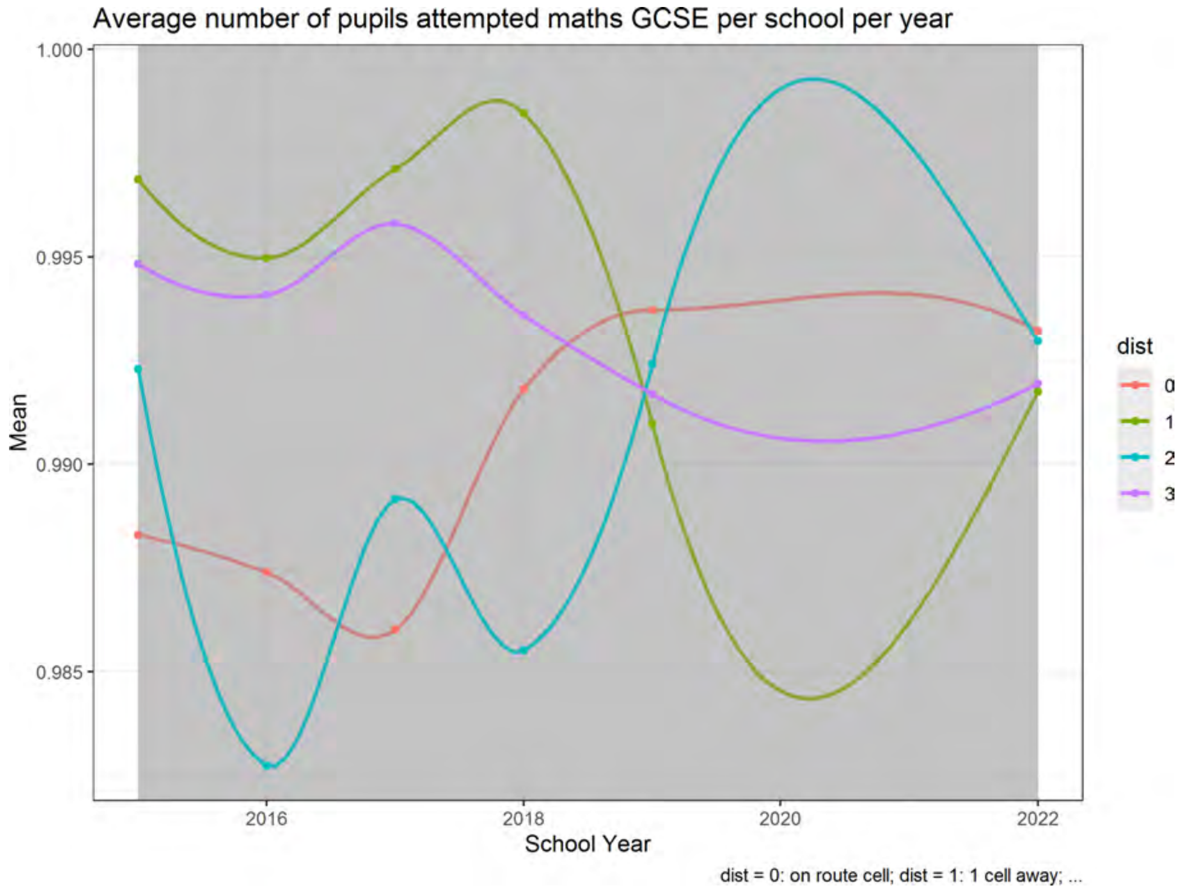
Figure A.2: Absence rate pre- and post-treatment



Notes: This figure shows the average absence rate in schools during autumn and spring terms of each academic year in each type of cell. 'dist' refers to the distance of a cell that the school is located on to the nearest Step Together route. Schools with dist=0 are located in cells on routes, while those with

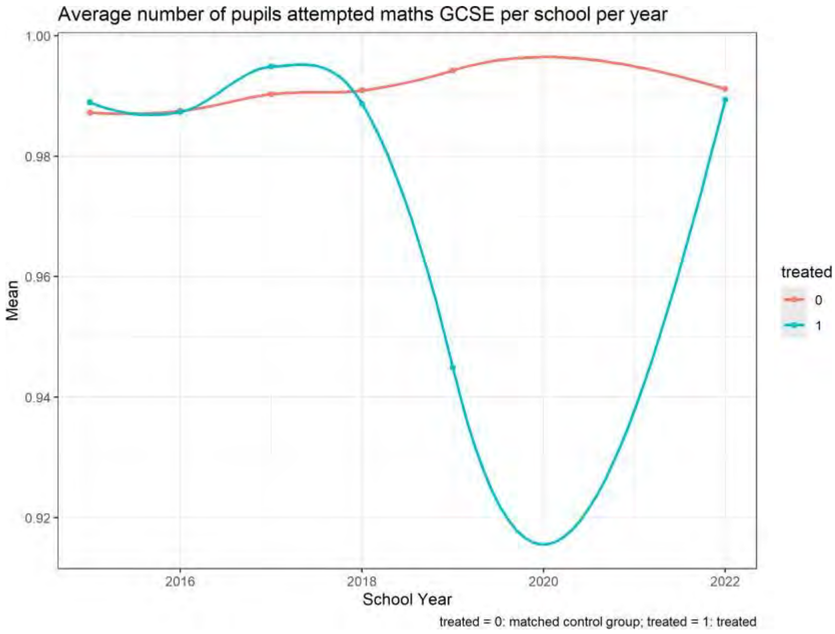
dist=1 are located on one cell away from a route (or equivalently, an on-route cell).
(Source: ONS)

Figure A.3: Percentage of pupils attempted Maths GCSE pre- and post-treatment



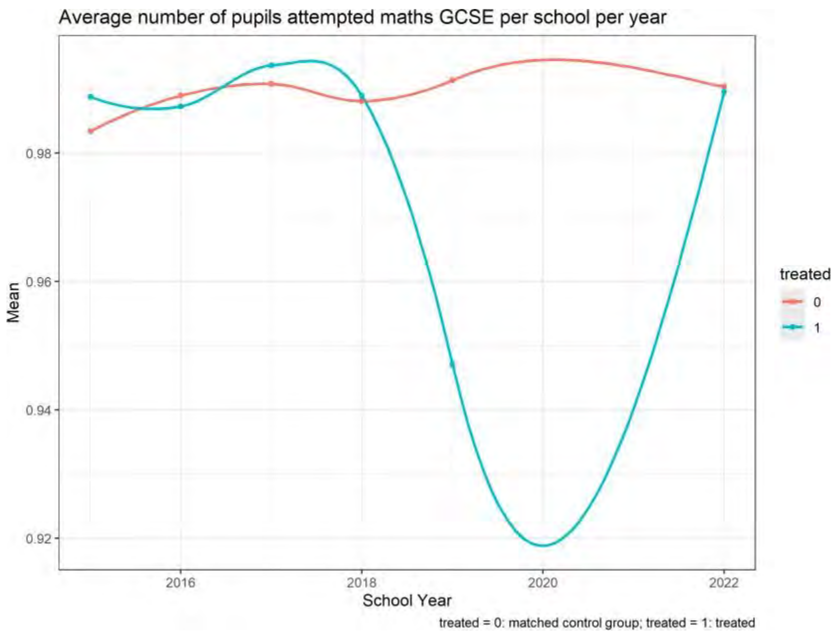
Notes: This figure shows Percentage of pupils attempted Maths GCSE in each school in each academic year in each type of cell. 'dist' refers to the distance of a cell that the school is located on to the nearest Step Together route. For example, schools with dist=0 are located in cells on routes, while those with dist=1 are located on one cell away from a route (or equivalently, an on-route cell). (Source: ONS)

Figure A.4: Percentage of pupils attempted Maths GCSE pre- and post-treatment



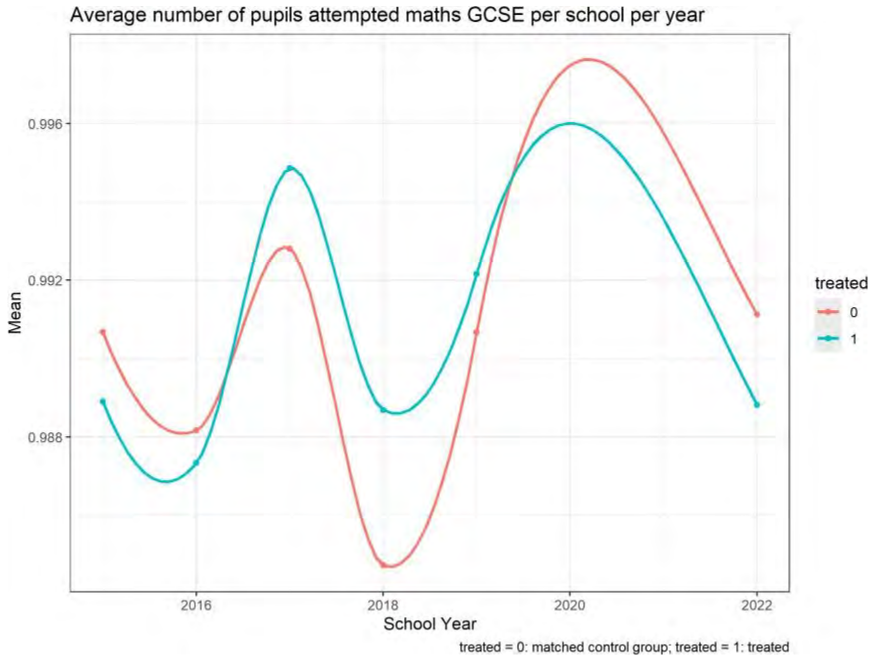
Notes: This figure shows the Percentage of pupils attempted Maths GCSE in schools in each academic year in treated vs matched control schools using 1-to-1 matching. (Source: ONS)

Figure A.5: Percentage of pupils attempted Maths GCSE pre- and post-treatment



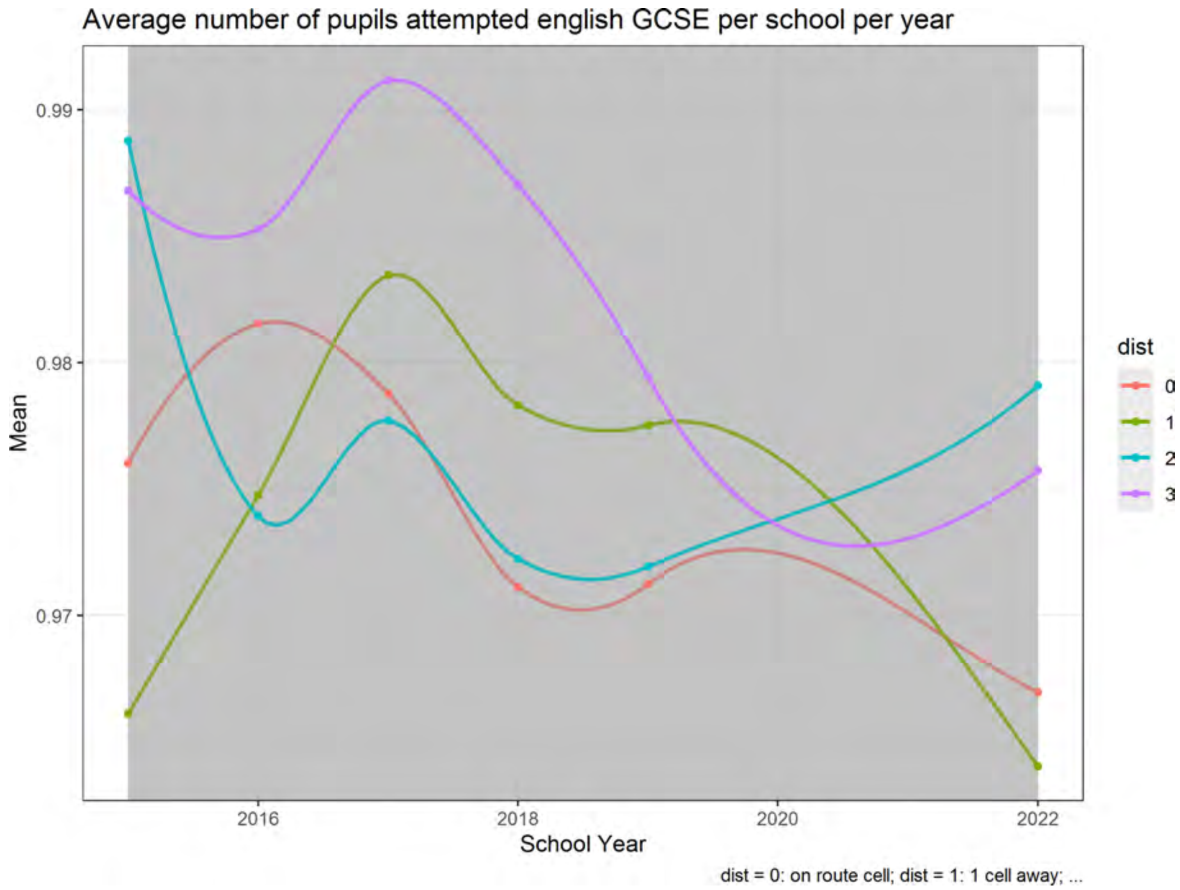
Notes: This figure shows the Percentage of pupils attempted Maths GCSE in schools in each academic year in treated vs matched control schools using k-nearest neighbour matching (k=3) matching. (Source: ONS)

Figure A.6: Percentage of pupils attempted Maths GCSE pre- and post-treatment



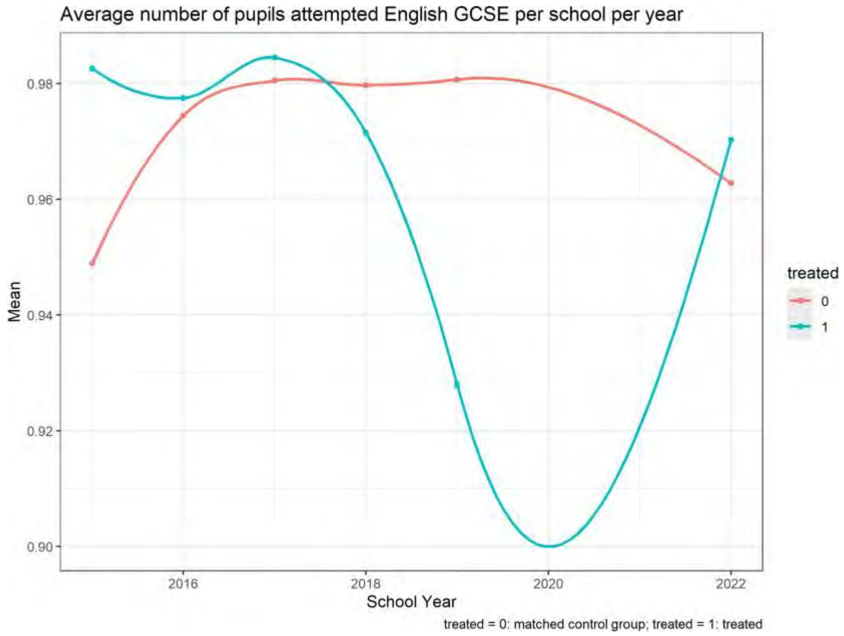
Notes: This figure shows the Percentage of pupils attempted Maths GCSE in schools in each academic year in treated vs matched control schools using Mahalanobis distance matching. (Source: ONS)

Figure A.7: Percentage of pupils attempted GCSE English pre- and post-treatment



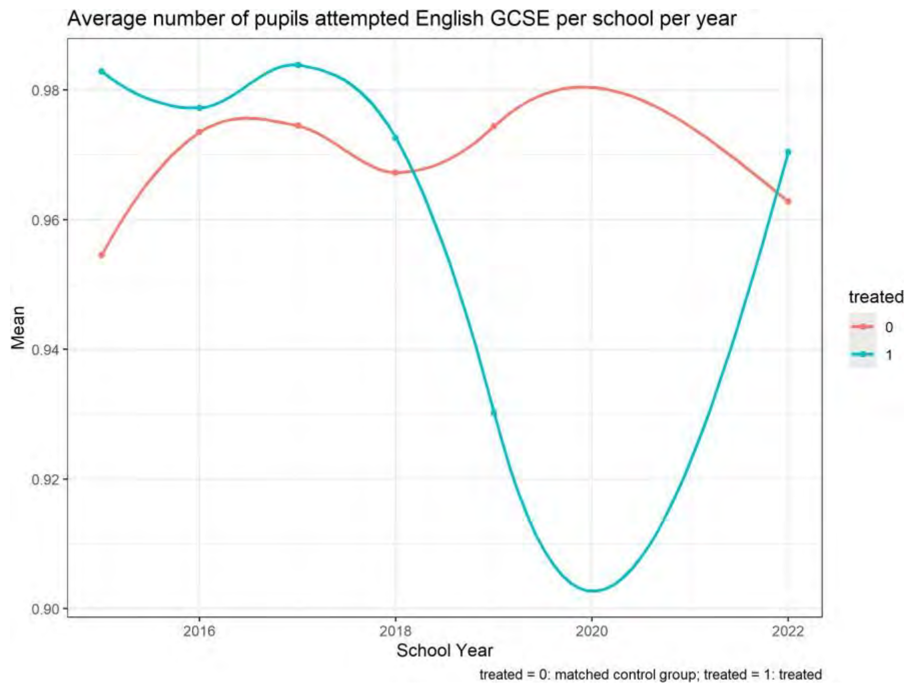
Notes: This figure shows Percentage of pupils attempted GCSE English in each school in each academic year in each type of cell. 'dist' refers to the distance of a cell that the school is located on to the nearest Step Together route. For example, schools with dist=0 are located in cells on routes, while those with dist=1 are located on one cell away from a route (or equivalently, an on-route cell). (Source: ONS)

Figure A.8: Percentage of pupils attempted GCSE English pre- and post-treatment



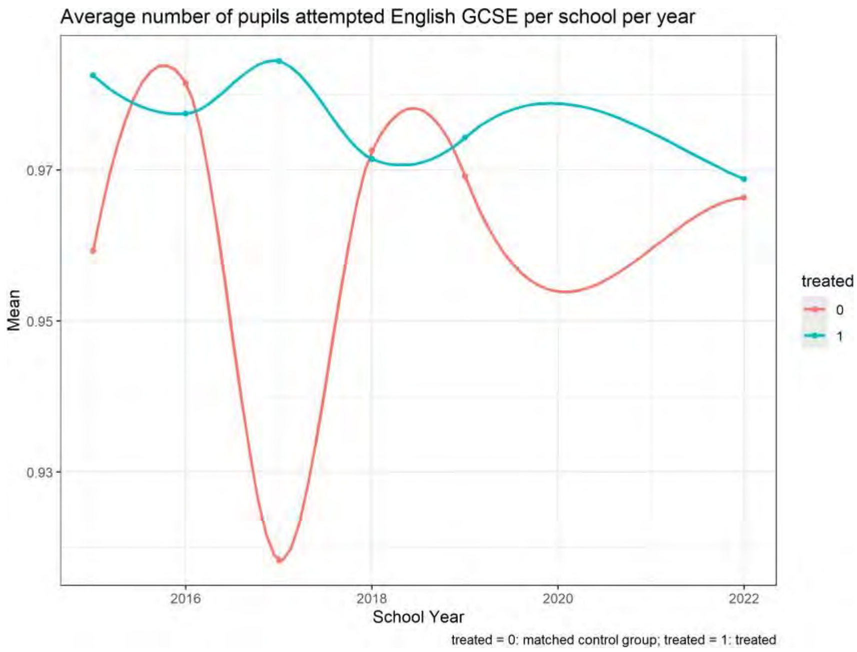
Notes: This figure shows the Percentage of pupils attempted GCSE English in schools in each academic year in treated vs matched control schools using 1-to-1 matching. (Source: ONS)

Figure A.9: Percentage of pupils attempted GCSE English pre- and post-treatment



Notes: This figure shows the Percentage of pupils attempted GCSE English in schools in each academic year in treated vs matched control schools using k-nearest neighbour matching (k=3) matching. (Source: ONS)

Figure A.10: Percentage of pupils attempted GCSE English pre- and post-treatment



Notes: This figure shows the Percentage of pupils attempted GCSE English in schools in each academic year in treated vs matched control schools using Mahalanobis distance matching. (Source: ONS)

Table A.1: Difference-in-differences regression output for absence rates

	Absence rate	
	Linear	Poisson
On-route × Post	0.001 (0.002)	-0.028 (1.018)
School FE	Yes	Yes
Year FE	Yes	Yes
Sample size	4,508	4,508

Notes: Outcome variables are absence rate in schools in each type of cell. On-route refers to schools located in cells through which a Step Together route runs. The control schools are located in cells more than three cells away from a treated cell. All standard errors are clustered by Step Together route.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (Source: ONS)

Table A.2: Difference-in-differences regression output for average standardised points in GCSE English

	Standardised point in English
	Linear
On-route × Post	0.221* (0.123)
School FE	Yes
Year FE	Yes
Sample size	3,864

Notes: Outcome variables are average standardised points in GCSE English in schools located in each type of cell. On-route refers to schools located in cells through which a Step Together route runs. The control schools are located in cells more than three cells away from a treated cell. All standard errors are clustered by Step Together route.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (Source: ONS)

Timeline

Table A.2 Timeline

Date	Activity
Aug'21- Oct'21	Inception Phase (LSE)
Sep'21	Evaluation Plan (LSE)
Oct'21- March'23	Data collection (LSE)
Jan'23-Aug'23	Analysis and Reporting (LSE)

Revised MDE calculations

Calculations for the minimum detectable effect (MDE) on crime and ASB previously provided in the study plan were based on assumptions which no longer reflect the actual empirical methodology used. In particular, the treatment and control units were assumed to be schools rather than grid squares currently used. This appendix provides revised MDE calculations for our main Poisson regression analyses.

Given that we use a DiD empirical approach (necessitated by non- random assignment to treatment), analytical calculation of statistical power and/or MDE is difficult. Instead, we use the simulation method described by Burlig, Preonas, & Woerman (2020), adapting the procedure to our context and data availability. This entails constructing synthetic treated outcomes data, by adding a treatment effect to representative data.

We use the five academic years of pre-intervention data (September 2014 to July 2019), since crime/ASB along the routes **prior** to the 2021–2022 academic year is likely to be similar to crime/ASB along the routes **during** the 2021–2022 academic year, if the Step Together pilot had not occurred. We also combine the on-route cells and one-away cells (i.e. cells directly next to routes), labelling these as potential treatment cells. One-away cells 10d to be similar to on- route cells, and adding these cells allows for richer simulated data, which improves the precision of the MDE estimates.

First, we label the first four academic years as pre-treatment and the last as post-treatment. Next, we randomly draw 104 treatment cells (the same number as actual on-route cells) from the 354 combined potential treatment cells to be labelled as on-route and therefore treatment cells. The remaining potential treatment cells are labelled as one-away cells. All other cells are kept unchanged. We then apply a given percentage treatment effect to the simulated post- treatment on-route cell outcomes, for example reducing violent crime by 10 per cent. Given this simulated data, we conduct the exact same Poisson regression procedures used to produce our main results, recording whether the estimated treatment effect coefficient is significant at the 5 per cent level.

By repeating these steps many times, the statistical power can be calculated as the proportion of repetitions in which the estimate is significant (i.e. the null hypothesis is rejected). To calculate the MDE given some effective power, we search for the treatment effect which approximately produces the desired power. Table A4 shows separately the MDEs calculated this way, for violent crime, property crime and ASB.

One drawback of using our pre-intervention data is that we can only assign a pre-treatment period of four years, rather than the five in our actual analysis. Given the simulation exploits fewer pre-treatment observations, our method underestimates power and overstates the MDE. In other words, our estimated MDE will be larger (in absolute terms) than the 'actual' MDE. To address this issue, we repeat our procedure with pre-treatment periods of 1,2 and 3 years to quantify the changes in MDE. The MDE remains almost identical across different pre- treatment period lengths, with the exception of ASB, where there is a small reduction towards 0 of the MDE. Hence, our estimates for ASB may overstate the actual MDE.

Table A.3: Estimated minimum detectable effect size

Outcome \ Effective Power	0.5	0.7	0.8
Violent crime	-0.0935	-0.113	-0.123
Property crime	-0.119	-0.140	-0.155
Antisocial behaviour	-0.124	-0.146	-0.159

Notes: Each cell shows the percentage reduction in crime/ASB required to achieve a given effective power, for a given offence type. For example, the upper-left cell means that to be able to detect a statistically significant effect on violent crime with probability 0.5, Step Together needs to reduce violent crime offences by 9.35 per cent in on-route cells. The calculations were performed given the actual numbers of treatment/control cells, with four pre-treatment and one post-treatment academic years of data.

Table A.4: Step Together Route Names, Start Date and Location

Route Name	LA	Start Date
Birmingham – Erdington	Birmingham	06/10/2021
Birmingham - Erdington High Street	Birmingham	13/12/2021
Birmingham - Jewellery Quarter	Birmingham	26/11/2021
Birmingham - Kingstanding	Birmingham	14/03/2022
Birmingham – Lozells	Birmingham	01/11/2021
Birmingham – Newtown	Birmingham	13/12/2021
Birmingham - Sparkbrook	Birmingham	22/11/2021
Birmingham – Sutton	Birmingham	13/12/2021
Coventry - Wood End	Coventry	28/02/2022
Dudley - Kates Hill	Dudley	13/12/2021
Dudley - St James	Dudley[20]	26/11/2021
Sandwell - Cape Hill	Sandwell	26/11/2021
Sandwell – Oldbury	Sandwell	13/12/2021
Solihull - Tudor Grange Park	Solihull	13/12/2021
Walsall - Palfrey (Joseph Leckie)	Walsall	13/12/2021
Walsall - Town Centre	Walsall	26/11/2021
Wolverhampton - Blakenhall	Wolverhampton	13/12/2021
Wolverhampton - City Centre	Wolverhampton	26/11/2021
Wolverhampton - Heath Town	Wolverhampton	13/12/2021

[20] Dudley route was suspended for a period of time. However, it was not possible to account for this issue in the current analysis due to the chosen method outlined in study plan. The DID methodology used to estimate the causal effects in the current framework does not allow turning off the treatment status in between.