Reporting Covid-affected evaluations

May 2021

Education Endowment Foundation

EEF evaluations have been affected by the Covid-19 pandemic in a variety of ways depending on the nature of the programme being assessed and the stage of the evaluation when the pandemic began. While there is no 'one size fits all' approach, to ensure that the influence of the pandemic on delivery and evaluation is consistently and transparently reported, and findings appropriately interpreted, we have developed a checklist of principles to guide report writing and reviewing.

Notes:

- This checklist is designed for impact evaluations. Some considerations relevant for other study types, such as pilots or scale-up evaluations, are not included here.
- Changes to the trial design and analysis approach made in response to Covid may or may be not have been pre-specified in updated protocols and statistical analysis plans (SAPs), depending on the speed and frequency with which changes were made. Reports should reference the most recent published versions of the protocol and SAP and make explicit where changes deviate from these.
- This checklist should be used alongside existing EEF guidance on statistical analysis, implementation and process evaluation (IPE), and cost evaluation. The structure follows the existing evaluation report template; Covid-related content should be included where appropriate.

Reporting checklist

Intervention

- □ The original intervention design, as well as any modifications necessitated by Covid, are clearly described. This could include:
  - Changes in **mode** of delivery – e.g. from in-person to online.
  - Changes in **timing, duration** or **frequency** of delivery – e.g. if it was cut short, interrupted, delayed, or extended in response to Covid.
  - Changes to programme **content** – e.g. refresher or 'top up' delivery introduced in response to Covid.
  - Changes to programme **recipients** or **deliverers** – e.g. due to reallocation of staff or 'bubbling' in schools.
  - Changes to the **aim** or **intended outcomes** of the programme – e.g. to provide additional school support during the pandemic.
- □ Where appropriate, the report presents an updated logic model to include Covid as a key contextual factor. If changes to the programme have been made in response to the pandemic, implementation logic underpinning the new intervention design is appropriately

defined, making explicit changes to the original logic model. Revised logic models reflect changes to key ingredients, causal mechanisms, mediators and moderators in a Covid-affected context.

Ethics and data protection

- □ The report confirms that changes to the intervention or evaluation design in response to Covid were submitted for ethical review as needed.
- □ If material changes were made to the data collection or data sharing arrangements for Covid-affected evaluations, the report includes updated MoUs, school and parent information sheets, and/or privacy notices provided to participants.

Evaluation design

- □ Any key changes to the original evaluation design in response to Covid are described – e.g. reducing the number of trial arms from three to two due to recruitment challenges.
- □ The report discusses the conditions under which schools participated in the evaluation. This includes:
    - Any **financial contribution** intervention settings made towards the programme in order to join the trial, and any adjustments to these in response to Covid.
    - Any **incentives** (financial or in-kind) offered to intervention and control participants to fulfil evaluation requirements, and the rationale for these. If the amount or timing of incentive payments changed in response to Covid, this is noted.
    - A description of the **control condition** (business as usual, waitlist control, active control with alternative intervention, etc.) and any Covid-induced changes.

Participant selection

- □ If changes to recruitment strategies or eligibility/ exclusion criteria were made in response to Covid (e.g. due to low recruitment numbers), these are clearly outlined.
- □ Where new cohorts or settings were recruited to a trial post-randomisation in response to Covid, e.g. due to re-running the experiment in the following school year, reports describe how new participants were recruited and any changes in the size or composition of the new study sample. Scenarios could include:
    - A trial tests a whole-class intervention for Y1 pupils. After cancellation of the Phonics Screening Check and other national assessments, delivery is repeated the following year with a **new cohort** of Y1 pupils from the original sample of schools, some of which may already have dropped out during the original trial year.
    - After a long period of interrupted delivery due to Covid, a whole-school intervention is re-run the following school year. Some of the original trial schools decline to continue in the trial so **new settings** are recruited to replace the losses.
    - Outcome testing for an intervention delivered in nursery is delayed to the next academic year, at which time children will be in reception, potentially at a different school. **New settings** are therefore recruited in order complete outcome assessment for the **existing cohort**.

- □ For quasi-experimental designs, the report discusses:
    - The hypothesised selection mechanism into intervention take-up and any expected changes in response to Covid – e.g. proactive targeting of settings with certain characteristics as part of Covid response.
    - The identification strategy used to create a comparison group, and any adaptations to the approach or plausibility of the assumptions in the context of Covid.

Outcome measures

- □ Changes to baseline and primary and secondary outcome measures necessitated by Covid are clearly and comprehensively described. These could include:
    - Changes to **administration** or **marking** – e.g. replacing independent invigilation with teacher administration or delivering remotely rather than face-to-face.
    - Changes to the **measure** used – e.g. substituting a previously-defined secondary outcome for the primary outcome if the primary outcome measure is no longer available, replacing in-school baseline testing with a historic NPD measure to reduce burden on schools, or revising the assessment administered to reflect changes in pupil age or learning expectations. Reports demonstrate that any new primary outcome measure satisfies EEF criteria, namely:
        - It is an attainment outcome;
        - It is a fair assessment of the intervention (in terms of the logic model and predictive validity);
        - It is a valid and reliable measure with proven measurement properties;
        - It is equally valid as the original primary outcome.
    - Changes to the **timing** of outcome testing – e.g. delaying due to partial school closures.
    - **Dropping** of pre-specified baseline or outcome measures – e.g. due to cancellation of national assessments or in-school testing. Analysis in the report is appropriate to the availability of outcome measures:
        - In cases where all primary and secondary outcomes are unavailable, the report does not include an impact evaluation (but explains the reasons for this).
        - If only secondary outcomes are available, and are not appropriate for consideration as primary outcomes (see above), presentation of impact analysis depends on the validity and relevance of the secondary outcome(s) as standalone measures of impact. For example, for an intervention focused on behaviour, a behaviour outcome would still be informative to report in the absence of attainment measures. In cases where a secondary outcome is not central the intervention logic model, was not intended to be interpreted independently of attainment data, or suffers from severe validity concerns, the report does not present secondary outcome analysis and explains the reasons for this.
- □ Potential biases or limitations associated with outcome measures, particularly relating to changes made in response to Covid, are transparently reported – e.g. if outcome

assessments were not designed for remote administration or deviate from prescribed procedures.

Power and sample size

- □ Sample size at the point of randomisation is reported for all trials, as well as any post-randomisation adjustments (see **Participant selection** above).
- □ Where changes to the evaluation design have been made in response to Covid, the minimum detectable effect size (MDES) at randomisation is recalculated to reflect changes to statistical uncertainty:
    - Where **new cohorts or settings** were recruited to the trial in response to Covid, MDES is recalculated to reflect the sample size at trial 'relaunch'. If (differential) attrition has already occurred prior to generating the 'new' sample, the report notes that the sample size in each group has been determined by something other than randomness and MDES estimates should be interpreted with caution.
    - Where the **primary outcome** has changed due to Covid, MDES at randomisation is recalculated to demonstrate the extent to which the new effect size is plausible for the given intervention and sample size.
    - Where the **baseline measure** has been lost due to Covid (e.g. due to cancellation of national assessments) or changed to something that has a lower correlation with the primary outcome, MDES at randomisation is recalculated accordingly.
- □ MDES at analysis is presented for reports that include an impact evaluation (see **Outcome measures** above).

Attrition

- □ Attrition is calculated on the basis of the sample size at trial (re)launch, while capturing any post-randomisation losses to the sample as appropriate. Scenarios may include:
    - Losses from the **original sample** if no changes or additions were made in response to Covid.
    - If a **new cohort** of pupils is recruited to re-run the experiment in response to Covid, pupil-level attrition is measured against the number of pupils in the new cohort. If setting-level attrition has occurred prior to recruitment of the new cohort, this is also included in the overall attrition calculation using the number of pupils in the respective settings in the original cohort.
    - If **new settings** are recruited post-randomisation to replace settings that previously dropped out, attrition is calculated on the basis of the number of settings in the 'refreshed' sample, provided there is no reason to suggest that attrition from the original sample affected balance between trial arms.
- □ Attrition is calculated as the ratio between pupils included in the primary analysis and those at trial (re)launch. For evaluations in which the primary outcome is no longer available (see **Outcome measures** above), pupil-level attrition need not be reported. However, information on the number of settings that withdrew from the evaluation is included for reference.
- □ Reasons for attrition/ loss to follow-up are provided.

Pupil and school characteristics

- □ Pupil and school characteristics are presented for the sample on which the experiment was run – that is, the original sample or a 'refreshed' one. If a 'refreshed' sample, differences in characteristics from the original sample at randomisation are noted.
- □ Balance is assessed between intervention and control groups for the sample on which the experiment was run.

Impact analysis

- □ Primary and secondary analysis models follow the form specified in the Statistical Analysis Guidance (2018), unless there is a very strong justification otherwise. One such justification could be where there is high non-adherence and differential loss to follow-up in outcome data across trial arms, leading data to be analysed non-experimentally (see *Change to a quasi-experiment* in Hedges and Tipton (2020), p.28). In this case, reporting standards for QEDs would apply.
- □ Sensitivity analyses or robustness checks are conducted to account for changes in sample characteristics or outcome measurement in response to Covid. Analyses not pre-specified in the SAP are clearly labelled and considered exploratory. Alternative specifications could include:
    - Controlling for characteristics on which treatment groups are not balanced;
    - Adding a dummy variable for testing mode (e.g. remote/ in-person) if mixed modes are used;
    - Adjusting for date of outcome assessment if the testing window extends longer than intended.
- □ Impact estimates from trials experiencing high levels of attrition are interpreted with appropriate caution.

Compliance and fidelity

- □ The report is explicit about what constitutes delivery 'as intended'. That is, intervention delivery may be assessed against the **original** delivery protocol to explore the extent to which recipients received the programme as designed, or against a **revised** delivery protocol that accounts for adaptations made in response to Covid. Where the latter differs from the original delivery protocol, changes are clearly described (see **Intervention** above).
- □ Compliance analysis is conducted on the basis of compliance indicators considered most appropriate by the evaluators in consultation with the developer – that is, the **original** indicators defined at the start of the trial, or a **revised** set of indicators developed during the trial to reflect changes to the programme in response to Covid. Where multiple compliance indicators have been developed (e.g. pre- and post-Covid), one is selected for the main compliance analysis and the other(s) included as sensitivity analysis.
- □ Where the primary outcome is no longer available (See **Outcome measures** above), compliance analysis is not conducted. However, where useful to report descriptively about compliance, e.g. the number of participants/ settings that satisfy compliance criteria, this is included.

FSM analysis

- □ Subgroup analysis of pupils eligible for free school meals (FSM) uses EVERFSM_6_P from the NPD.
- □ Changes in the proportion of FSM pupils relative to assumptions, e.g. due to an increase in FSM pupils during the pandemic, are noted in the text and in the MDES at analysis.

Implementation and process evaluation

- □ The report clearly describes any changes to the IPE design or data collection methods in response to Covid. These could include:
    - **Cancellation** or **reduction in scope** of planned activities – e.g. due to partial school closures, inability to visit settings, or to reduce burden on schools.
    - **Addition** or **expansion** of IPE activities – e.g. exploring implementation in the context of Covid. Where **new research questions** were added in response to the pandemic, these are noted in the report.
    - Changes to the **data collection methods** used – e.g. replacing lesson observations with document analysis of lesson plans or focus groups with telephone interviews.
    - Changes to the **timing** of activities – e.g. delaying data collection due to partial school closures.
- □ Participation numbers and/or response rates are provided for all IPE activities and findings interpreted appropriately. Where efforts to encourage participation were curtailed due to Covid (e.g. by sending fewer reminder emails) or changes made to sampling procedures to facilitate participation (e.g. by expanding the criteria for case study selection) this is described.

Business as usual

- □ The report describes the nature of instruction in comparison schools in both pre-Covid and Covid-affected periods, if applicable, and attempts to capture the range of provision during the project duration.
- □ Participation of trial pupils/ settings (intervention and control) in national Covid recovery programmes, e.g. National Tutoring Programme, Nuffield Early Language Intervention (NELI), or other interventions taken up in response to Covid, is described.
- □ Where (sufficient) information on business as usual or concurrent interventions is not available, this is clearly noted as a limitation.

Cost evaluation

- □ Costs presented reflect actual programme costs incurred by relevant stakeholders. If these differ from 'normal' costs outside the Covid context, the latter can also be presented but are not used for the headline estimate (see Principle 2 in the cost evaluation guidance (2019)).

- Trials with no primary impact estimate are not assigned a cost rating so need not include a cost evaluation. However, where complete cost data has already been collected, a unit cost is presented for reference.

Conclusion

- The report comments on the extent to which the results of the evaluation support the original or revised logic model as appropriate.
- The report reflects on the potential for misattribution of causal agents in the Covid context – e.g. a well-packaged online intervention outperforming rapidly prepared comparison instruction because it is better online instruction, not better instruction overall (see *Construct validity of cause* in Hedges and Tipton (2020) p.31).
- Difficulties in the interpretation of impact estimates – e.g. due to changes in programme delivery or business as usual, loss of baseline or outcome measures, attrition, etc. – are transparently acknowledged.
- The report discusses the extent to which any changes made to intervention delivery in response to Covid will be retained as part of future implementation (if known).

References

Education Endowment Foundation, 2019. *Cost evaluation guidance for EEF evaluators*, London: Education Endowment Foundation. Available from: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/Cost_Evaluation_Guidance_2019.12.11.pdf

Education Endowment Foundation, 2019. *Implementation and process evaluation guidance for EEF evaluations*, London: Education Endowment Foundation. Available from: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_guidance.pdf

Education Endowment Foundation, 2018. *Statistical analysis guidance for EEF evaluations*, London: Education Endowment Foundation. Available from: https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

Hedges, L and Tipton, E. 2020. Addressing the Challenges to Educational Research Posed by Covid-19. Northwestern Institute for Policy Studies Working Paper Series WP-20-47. Available from: https://www.ipr.northwestern.edu/our-work/working-papers/2020/wp-20-47.html