Foundation

Global Safety Evidence Centre

Research Methods Guide: Data collection, analysis and trials principles for the safety space

Charity Impact

Tools & Methods

Resource

About Lloyd's Register Foundation Global Safety Evidence centre

The Lloyd's Register Foundation Global Safety Evidence Centre is a hub for anyone who needs to know 'what works' to make people safer. The Centre collates, creates and communicates the best available safety evidence from the Foundation, our partners and other sources on both the nature and scale of global safety challenges, and what works to address them. It works with partners to identify and fill gaps in the evidence, and to use the evidence for action.

To find out more about the Global Safety Evidence Centre, visit *gsec.lrfoundation.org.uk*

About Lloyd's Register Foundation

Lloyd's Register Foundation is an independent global safety charity that supports research, innovation, and education to make the world a safer place. Its mission is to use the best evidence and insight to help the global community focus on tackling the world's most pressing safety and risk challenges.

To find out more about Lloyd's Register Foundation, visit *lrfoundation.org.uk*

Lloyd's Register Foundation, 71 Fenchurch Street, London, EC3M 4BS, United Kingdom

Lloyd's Register Foundation is a Registered Charity (Reg. no. 1145988) and limited company. (Reg. no. 7905861) registered in England and Wales, and owner of Lloyd's Register Group Limited.

Copyright © Lloyd's Register Foundation, 2026.

This work is licensed under *CC BY-SA 4.0*

doi.org/10.60743/0QED-1J27

Contents

Glossary of terms 3

Chapter 1. Introduction: The Global Safety Evidence Centre and primary evidence in the safety space 4

Lloyd's Register Foundation - The Global Safety Evidence Centre 4
A 'what works' approach 4
This guide 5
Movement towards standardisation of research for social policy 6
Challenges for primary safety research 6
Further reading 6

Chapter 2. Primary evidence research foundations and principles 7

Importance of good methodology 7
General quantitative research principles 7
Design with methods that suit the evidence base 8
Design in consideration of the statements you are hoping to make 9
Solid research questions 9
Defined objectives, metrics and outcomes 9
Peer review consideration 10
Pre-registration and transparent documentation 11
Further reading 12

Chapter 3. Ethical principles and considerations for data collection 13

Why ethics matter in primary research 13
Ethical principles for safety research 13
Ethics review and oversight recommendations 14
Further reading 15

Chapter 4. Data collection tools, validation and data types 16

Tools for safety primary evidence research 16
Validation of surveys and questionnaires 16
Self-report versus observer-rated data 16
Validation of interviews and qualitative data 18
Further reading 18

Chapter 5. Quantitative data characteristics, distributions, probability and statistics 19

Distributions 19
Probability – Frequentist and Bayesian methods 20
Statistical analysis – an introduction 20
Significance 20
Multiple comparisons 21
Confidence intervals 21
Power 21
Further reading 21

Chapter 6. Case studies, observational and quasi-experimental research 22

Case studies 22
Participatory Action Research (PAR) 22
Observational research 22
Causality in research 23
Quasi-experimental studies 23
Further reading 23

Chapter 7. Experimental research and trials 24

Foreword: need for experimental research in the safety space 24
Randomisation Theory 24
Trial types and principles 25
Ethical considerations 27
Effectiveness and efficiency 29
Further reading 29

Chapter 8. Data management, storage, visualisation and reporting 30

Pre-statement of hypotheses and analysis plans 30
Open and reproducible analysis tooling 30
Open data sharing 30
Long-term storage 31
Visualisation and communication of findings 31
Multiple levels of description for different audiences 32
Further reading 32

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Glossary of terms

Term	Description
Evidence	information that supports or challenges a conclusion, statement, or belief. It includes quantitative (nominal - numbers or statistics) and/or qualitative (non-numerical, words, stories or observations) information. Evidence should be systematically and ethically gathered to be reliable.
Research	a defined and planned approach to collecting and analysing data in order to answer specific research questions.
Data	information and statistics for reference or analysis. Data can be qualitative or quantitative. There are various data sources, including monitoring data to track service performance, primary data for research or evaluation, and secondary data collected by another organisation or for another purpose.
Insight	the understanding gained from analysing data, identifying patterns, and drawing meaningful conclusions. Insights go beyond the basic findings to provide interpretation and analysis, identifying the 'so what' behind the data.
Rigour	the application of systematic research methodology to achieve robust, standardised and reproducible results.
Robustness	durability and repeatability of findings provided by rigorous methodology.
Credibility	the accuracy and usefulness of research methods and findings.
Independent variable	that which changes or is changed in a study.
Dependent variable or outcome variable	that which is measured or observed as a result of a changes in independent variable.
Primary evidence	includes data collection methods, data collected first-hand and analysis of that data.
Secondary evidence	includes analysis and synthesis of pre-existing data and information.
Research question	the research enquiry - what the research will answer.
Hypothesis	a specific prediction on the effect or outcome expected and its direction in experimental research.
Null hypothesis	a baseline prediction for an absence or opposite effect (to the hypothesis).
Data distribution	the characteristics of the dataset or variable.
Probability distribution	the characteristics of the probability function which the data is tested against in statistical analyses.
Conditions	different groups or levels of interest in a study.

Term	Description
Test group	a study group that has characteristics of interest (terminology used in observational studies).
Treatment group	a study group of interest, where there is an intervention or manipulation.
Control groups	a study group where there is not an intervention or manipulation (or without characteristics of interest in observational studies) which is used for comparison.
Intervention	that which is changed within a treatment group; a manipulation, activity or programme of activities.
Observational studies	concerned with identifying patterns, trends and associations in without a causal framework.
Experimental studies	concerned with understanding the effect of independent variables (changes or interventions) on outcomes in a causal framework.
Prospective research	has a predictive or causal framework – usually intervention studies, where outcomes are explored in relation to planned changes.
Retrospective research	looks at data that exists for patterns and relationships, where outcomes are explored in relation to unplanned or existing changes.
Effectiveness	a cornerstone of experimental research and showing what interventions work. In experimental research, the term effectiveness has a specific meaning, relating to data collected in a causal framework where statistical differences in outcomes are shown.
Efficiency	a cornerstone of research evaluation. In experimental research, efficiency relates to an assessment of resources required for effective interventions.
Safety	the prevention of accidental death, serious injury (human safety), loss or damage (of/to physical assets or the environment). The control of hazards to an acceptable level of risk. It is a public good that is produced and is an emergent property of systems. We take a systems-based, evidence-led approach to safety that aligns with the complex realities of lives and economies. We expand the definition of safety beyond occupational health and safety to encompass the full interaction between people, infrastructure, and the environment.
Safety Science	knowledge about safety and risk related issues more broadly, and how to assess and manage them. This covers - Knowledge about safety related issues; Tools and methods for safety such as the development of concepts, theories, principles and methods to understand, assess, communicate and manage (in a broad sense) safety; The discipline as a whole: covering the totality of relevant educational programmes, journals, papers, researchers, research groups and societies.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Chapter 1. Introduction: The Global Safety Evidence Centre and primary evidence in the safety space

Authors – Emily Lambe & Nancy Hey

Evidence is critical to improving safety – but often does not yet exist or is not easily accessible. Across the world and across different sectors, there is a huge opportunity to improve safety outcomes by generating better quality evidence on both the scale and nature of the challenges, and on what works to address them. But simply generating that evidence is not enough – it must also be relevant, understandable, accessible and actionable by those in a position to put it into practice (*Engineering a safer world: Lloyd's Register Foundation Strategy 2024-2029*).

Lloyd's Register Foundation - The Global Safety Evidence Centre

Lloyd's Register Foundation is a global safety charity with a mission to engineer a safer world. We do this by harnessing our heritage to shape a safer and more sustainable ocean economy for the future and finding and sharing the best evidence and insight on what works to improve safety. The Foundation's Global Safety Evidence Centre (GSEC) was established in 2025 as a comprehensive hub for anyone who needs to know 'what works' to make people safer. The Global Safety Evidence Centre works with a wide range of institutions, teams and practitioners and with global practitioner bodies and international organisations to ensure that the most important research questions are answered in the best ways and in a timely manner.

The Centre's audiences include:

- Practitioners: anyone who can use our evidence in their work to support safe work, including employers, managers and supervisors, HR professionals, regulators, policy makers and campaigners. Practitioners may work in industry or government, public or private sectors, charities and communities.
- Researchers: anyone who generates evidence which can be used by others in their work, including university academics, evaluators and research consultants.
- Lloyd's Register Foundation: teams around the Foundation use our evidence to make decisions across our portfolios and provide strategic funding.

The Evidence Centre operates on 'Collate, Create, Communicate' functions of evidence generation, being independent, evidence informed, collaborative, open, iterative and practical. As a global Centre our aim is to gather, generate and provide evidence internationally, which is both driven by and supported by initiatives like our *Ocean Centre* partnerships in Brazil, Ghana, Kenya, Bangladesh, India, Indonesia, Philippines.

As a trusted global source of evidence on safety, we support research which is:

1. Robust and credible: using tested methods, established standards and transparent reporting so that practitioners and researchers can be confident in our findings.
2. Relevant and useful: responding to the evolving needs of practitioners and generating insights they can act on in different contexts around the world.
3. Communicated well: through simple and accessible means to reach as many people as possible and help them put the evidence into practice. We work closely alongside researchers and bring together practitioners and policy makers to make sure any evidence we publish in our Global Safety Evidence Library meets these criteria.

A 'what works' approach

'What works' is a method that can be used to improve the impact that research findings have on people's lives and the effectiveness of the use of time, money and effort towards safety of life, property and livelihoods. It is based on the principle that good decision making is underpinned by good evidence, and if that evidence isn't available, robust ways of generating that evidence should be established. 'What works' recognises that research evidence on its own is not enough; you need to know how and why something works, for who, and finally, how to implement what is known. (Lloyd's Register Foundation, 2024)

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

This guide

This resource accompanies the Foundation's *Evidence Review and Synthesis Methods Guide*. The aim of this Research Methods Guide is to provide practical guidance on research methods, data collection and analyses principles, and trial approaches for safety research. It is a foundational resource within the Evidence Centre's 'Create' stream, designed as a tool for researchers, practitioners and policy makers in this space globally. The research techniques in this guide are well established, however our aim for this guide is for it to be practical, accessible and applicable to safety research in different contexts by different organisations.

Research design and implementation involve a series of decisions. This guide includes concepts that researchers should have an awareness of in order to make those research decisions.

What this guide is:

- A guide on concepts and key research principles, which are essential for quality research production, and the standards we align to.
- Provides research design options and signposts to resources for further exploration.
- A levelling document for organisations across the safety space who are considering research and are looking for guidance and inspiration.
- An expectation document to facilitate understanding of the research standards for funding with the Global Safety Evidence Centre.
- How we might plan a research project, documented for transparency.
- For teaching and helping people avoid small common errors.
- Currently focused on data collection, quantitative methods and trials.
- Currently focused on organisational and social research, though applicable to different contexts such as technology and environment.

Researchers and topic specialists often spend a lifetime studying research methodology and that cannot be replicated in this guide. Research design decisions must be made in consideration of the context, research questions and expected data/outcomes, so there are no one-size-fits-all research initiatives.

What this guide is **not**:

- An exhaustive guide on how to do research.
- Inclusive of field-specific methods (e.g. engineering, economics, data science & AI).
- A statistical advisory document.
- A subjective, qualitative, appraisal and evaluation methods document (yet).
- A guide on context specific safety data sources.

It is acknowledged that many safety data sources are secondary, i.e., collected through existing industrial processes. Many of the same principles apply to both primary and secondary data analysis. Secondary data analysis tends to be observational which is covered to an extent in this guide. However, the main focus of the guide is primary methods in order to support the collection of quality data and applications for experimental/trial research projects currently.

It is also worth saying that there are active and well-developed field-specific evidence practices (technical, infrastructure and engineering) that are not covered here but are under consideration for future guides in different parts of Lloyds Register Foundation (e.g. *AI assurance, Data-centric engineering*).

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Movement towards standardisation of research for social policy

Standardisation involves the generation of consistent processes and documentation, universal methods and defined metrics and measures, to achieve research consistency and knowledge transferability. Quantitative methods (based on numbers and statistics) are synonymous with standardisation, as the goal is to use common tools, metrics, methods and analysis to build comparable findings. With huge and growing data capabilities across the world, standardisation is increasingly important for social policy. Part of the work at the Global Safety Evidence Centre – through the Safety Science programme – is to develop these tools. This guide supports the Safety Science programme.

Challenges for primary safety research

Conducting primary evidence projects and collecting good data in the safety space is inherently challenging for several reasons:

1. The concept of safety spans all work types and a huge number of disciplines, which makes it difficult to generate universal approaches and comparable datasets.
2. The environmental and societal context of safety at work is extremely variable and it is important to note that standardisation cannot mean one-size-fits-all solutions.
3. Relatedly, methods employed will differ significantly; an RCT in western manufacturing will have different considerations to an iterative ethnographic small scale fishing study in Asia.
4. Research must adhere to the ethical principle of no harm, meaning interventions in the safety space are limited and should be carefully designed.
5. A consensus on safety tools, assessment, indicators and metrics for best practice across different contexts is not yet established.
6. Safety research might not be immediately rewarding. Unlike drug trials, for example, where the treatment conditions are simply defined, i.e., drug-placebo, effective safety interventions require a broad view and are likely to involve combined and cumulative factors (akin to a therapy model). Finding the best combinations for different contexts is the goal.

Learning can be taken from primary evidence processes and resources in the medical field to help with this goal. The international standard which all clinical trials must follow, Good Clinical Practice (ICH-GCP), for example, is in place to ensure participant rights, safety and data integrity in research. While no equivalent regulatory or procedural system exists for safety research overall (though various important standards exist, e.g., International Organization for Standardization [ISO 45001], 2018; International Labour Organization [ILO], 2001), in the following chapters transferable ideas and examples from healthcare, social science, engineering and environmental science will be used to explore research principles for safety.

At the Global Safety Evidence Centre, we take a systems-based, evidence-led approach to safety that aligns with the complex realities of lives and economies. **We expand the definition of safety beyond occupational health and safety to encompass the full interaction between people, infrastructure, and the environment.** It is hoped that this guide is useful in facilitating quality research in this space.

Further reading

Lloyd's Register Foundation. (2024). Strategy 2024–2029. https://www.lrfoundation.org.uk/about-us/strategy

Lloyd's Register Foundation. (2025). Global Safety Evidence Centre. https://www.lrfoundation.org.uk/gsec

Breckon, J. (2017). Using research evidence: A practice guide. Nesta. https://www.alliance4usefulevidence.org/publication/using-research-evidence-a-practice-guide/

The Alan Turing Institute. (2026, April 28). Data-centric engineering. https://www.turing.ac.uk/research/research-programmes/data-centric-engineering

University of York. (n.d.). Centre for Assuring Autonomy. https://www.york.ac.uk/assuring-autonomy/

International Labour Organization. (2001). Guidelines on occupational safety and health management systems: ILO-OSH 2001. https://www.ilo.org/publications/guidelines-occupational-safety-and-health-management-systems-ilo-osh-2001

International Organization for Standardization. (2018). ISO 45001:2018 occupational health and safety management systems—Requirements with guidance for use. https://www.iso.org/standard/63787.html

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Chapter 2. Primary evidence research foundations and principles

Author: Emily Lambe

Chapter 2 describes research principles and concepts that GSEC would expect to see evidence of, or awareness of, in strong proposals. Sometimes details are excluded because applications are not written by research specialists, other times principles are overlooked, but the aim of the chapter is to support quality funding applications and research plans.

Primary evidence research involves the first-hand collection of quantitative and qualitative data. These data types have distinct principles and associated research methods. Quantitative methods are concerned with objective, nominal data, whereas qualitative methods are concerned with subjective, experience-rich, non-numerical data. Both are important for our understanding of safety. However, this version of the guide will focus on quantitative methods.

Importance of good methodology

A core aim of quantitative methods is to control certain systemic and individual phenomena that can dilute the quality, meaningfulness and usefulness of the evidence pool.

- Bias is one phenomenon and comes in many forms; selection, measurement, researcher, confirmation, reporting, attrition and analysis bias, as well as participant bias. Bias should be considered anywhere where there are analytical choices made, i.e., researcher degrees of freedom. Two well-known examples of bias are hypothesising after the results are known, HARKing, and p-hacking – manipulating analyses, and/or cherry-picking outcomes, in order to report statistically significant results. Often, pressure to publish in journals that prefer significant results, and desire to find a 'true' or impactful effects can drive these actions.

Importantly, HARKing applies to outcomes reported as if they were predicted but were not. It does not prevent exploratory data analysis. However, the main point is that exploratory analyses should be planned and reported as such.

- Random error, otherwise known as noise, is another phenomenon. Random error can affect the precision (or similarly repeatability) of results. Random error can be addressed with increased sample size and, though important, is less difficult to mitigate than bias in this respect.

General quantitative research principles

To minimise confounding factors, primary evidence in the safety space should be developed in alignment with core quantitative research principles. These principles are outlined below and accompanied by hypothetical safety examples.

Table 1. General quantitative research principles

Principle	Why Important	Safety Example
Objectivity – data should be collected, analysed and interpreted free from bias.	To avoid skewed and misrepresentative data.	Reporting all correlations between safety indicators, not only those statistically linked to reduced injury rates.
Measurability – observations or changes must be clearly defined and numerically measurable.	Ensures tools and measures genuinely assess what is intended.	Using a safety culture questionnaire that measures relevant components (e.g., leadership), validated through pilot testing or prior data.
Representativeness – data should reflect the population or phenomenon of interest.	Ensures the sample reflects the group you want to draw conclusions about.	Stratifying World Risk Poll data collection by age to analyse or control for age-related factors.
Internal validity – confidence that relationships between variables are real, not due to confounders.	Ensures outcomes are due to the independent variable in experimental designs.	Randomising participants in a safety training RCT to control for spurious group differences.
Generalisability (external validity) – whether findings would hold in other populations, settings or times.	Ensures findings apply to largescale, real-world situations.	A five-thousand-person national survey is more generalisable for UK safety trends than 300 surveys from one region.
R1. Repeatability – same team, same methods, same conditions → same results.	Builds confidence that findings are stable under identical conditions.	Readministering a safety climate survey to the same workers under the same conditions and obtaining similar scores.
R2. Reproducibility – different team, same methods → same results.	Demonstrates that findings are not dependent on a single research group.	Another research team using the same safety training protocol and achieving similar improvements.
R3. Replicability – different team, different methods → same conclusions.	Shows that findings hold across varied approaches, strengthening evidence.	Independent teams using different methodologies to study PPE effectiveness and reaching similar conclusions.
Transparency and accuracy – data collection, processing and analysis should be precisely recorded and openly available.	Encourages repetition, reproduction and replication; discourages misconduct.	Preregistering a safety study protocol or analysis plan in the Open Science Framework repository
Reliability – overall consistency and stability of measures and outcomes (supported by the 3 Rs).	A cornerstone of quantitative research that underpins trustworthiness.	Findings (patterns and effects) or measures supported by the 3Rs.

Note: the 'R's' have slightly different definitions across disciplines, however when considering safety research as a whole, the above can be applied.

Adherence to these research principles supports standardisation in the field and helps ensure robust, systematic, high-quality, rigorous and transparent quantitative research. The aim is more good data not necessarily just more data!

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Design with methods that suit the evidence base

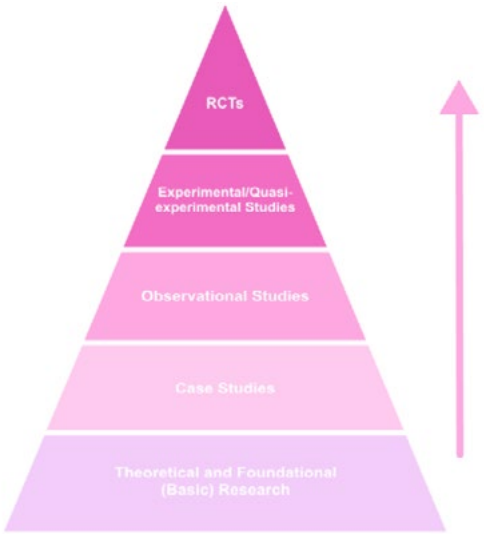
Before primary data is collected, an assessment of the evidence base should be carried out and a strong rationale for the research methods, based on existing (or lack of) research, generated. If little research exists in an any area, the evidence base would first benefit from qualitative insights, case studies and observational studies (and review/synthesis methods where evidence exists).

Studies with more robust experimental designs, should be well-grounded in previous research. This systematic progression of the evidence base helps prevent inaccurate findings being taken as concrete. For example, an RCT study can be robust in terms of methodology but if not grounded in previous research, results can still lack generalisability. In fact, an RCT without a solid foundation can cause more harm than good for the evidence base because of the general assumption that RCTs are strong, robust and therefore accurate.

Evidence synthesis work can (and should) be used to both identify research gaps, develop primary evidence research questions and identify appropriate methods, as well as summarising primary evidence that already exists.

Broadly, primary evidence methods include case studies, observational studies, quasi-experimental studies, experimental studies and trials and Randomised Controlled Trials (RCTs).

Figure 1. Pyramid of primary evidence generation



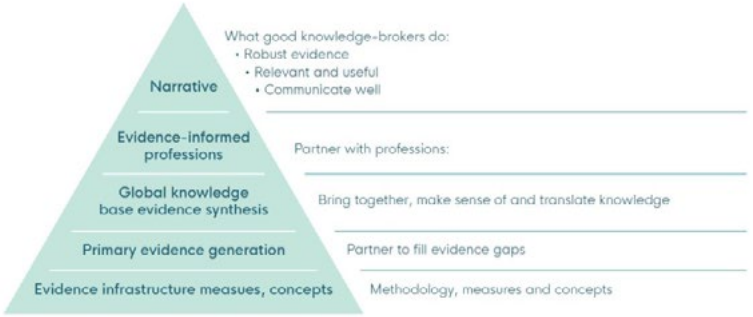
The purpose of figure 1 is to show that primary evidence should have a foundation. Case studies should have an existing theoretical foundation, for example, and intervention (experimental) studies should be grounded in observational findings.

That said research is an iterative process. For example, case studies can drive theory and further foundational research. But, the pyramid illustrates the idea that if no evidence exists in the level below, a study within a given level would be scientifically, and potentially ethically, sub-optimal at best.

Importantly this is not a schematic that can cover all research in the safety space. For example, in engineering, the best evidence pathway is question-led (controlled testing, failure analysis, verification, validation or validated simulation) which are not progressions toward an RCT.

Furthermore, not everyone will have the infrastructure or financial resources or skills to run an RCT, so the pyramid should be considered in the context of doing the best possible research for the context. If a case study only is possible or relevant, do the best one possible planned with consideration of the endpoints and conclusions.

Figure 2. Pyramid of evidence generation and uses



The purpose of figure 2 is to show how primary evidence supports the wider 'what works' framework for evidence-based knowledge generation and practice.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Design in consideration of the statements you are hoping to make

Another fundamental part of research method selection is anticipating the data that will be generated and what conclusions can be made with that data. Key to this are:

- Research questions – open ended topic enquiry - what is the research asking? All research projects (quant & qual) should start with research questions.
- Hypotheses – specific predictions based on the direction of the effect (research result) that is expected. Not applicable to all research but required for quantitative experimental research.

What conclusions (broadly) can be made:

- Descriptive – what patterns exist in dataset D
- Association based – does X relate to Y (observational)
- Comparative – what differences exist between A B C
- Causal – do changes in X lead to changes in Y

In human and organisational and social research, only with an experimental design, ideally an RCT, can causal statements about an intervention i.e. safety intervention X improved safety outcome Y, be made. This is further explored in chapter 7.

Solid research questions

Characteristics of a strong research question include:

Specific and clear

For example, 'is increased communication between management and employees associated with reduced accident and injury rates?' is more specific than 'does management communication affect employee safety?'

Note: a broader research question is fine if the analyses are exploratory, but the study documentation should explicitly state the exploratory nature of the research question and analyses.

Neutral, not leading

In the example 'how do leading indicators improve safety?', there is an assumption that leading indicators do improve safety. Unless that statement is known and accepted through previous research, it is a leading question. 'Are X leading indicators associated with reduced accident rates', for example, would be a better start.

Feasible

It seems obvious, but the research question should be something that can be answered. 'What is the best way to improve safety outcomes', for example, is an interesting question. However, it is not feasible to investigate the full spectrum of safety variables in one primary evidence project. Other techniques such as evidence synthesis would be more appropriate for that question.

Defined objectives, metrics and outcomes

Having defined inputs and outcomes at project initiation is crucial in primary research for several reasons; they allow measurability, tool selection and relevant design choices, reduce bias and possible cherry-picking practices and support ethical research as participants should know upfront what data will be collected and how. Overall, the integrity of data collection cannot be assessed without pre-defined inputs and outcomes determined at the initial stage of any project.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Peer review consideration

Peer review is a vital quality control mechanism and helps to ensure scientific and ethical integrity in primary research. At its core, it involves an assessment of project aims, ethics, research methods and interpretation, by independent experts to determine whether the study meets the quality and validity required. Peer review of key research documentation should occur at distinct stages of the project life cycle; usually at the planning and design stage, the interim period (and with any study changes), the analysis stage and following write up.

Points covered in primary evidence peer review meetings can include:

- Ethical considerations such as informed consent and the principle of no harm (see Chapter 3)
- Suitability of the design for the research questions
- Population selection and characteristics
- Randomisation methods
- Suitability of the analyses for the design and data

Peer review helps identify errors or any gaps in reasoning that mean the aim of the project cannot be achieved in a rigorous and ethical way. There are no specific standards for peer review in the safety research space, however, good practice is well documented across academia and other disciplines from which learning can be taken (*Kelly, 2014*).

Peer review in healthcare research

In the medical field where safety is critical and findings are expected to impact healthcare and public health policy, peer review standards and steps are integral and mandatory to the research process. The main reason is that medical research is highly regulated - research documents undergo several stages of peer review in order to be legally and operationally approved. In the UK, the bodies responsible for clinical trial approval are The Medical and Healthcare products Regulatory Agency (MHRA) function of Government and designated local Research Ethics Committees (REC). In this process, before a study commences the study protocol, analysis plan, participant facing documents, drug- or device-specific information and research team profiles, at a minimum are reviewed by more than one external expert *>> UK clinical trial regulations*. The same principles apply to the safety space, particularly for experimental trials projects.

Peer review in safety research

Arguably all projects assessing safety outcomes should be subject to peer review to ensure freedom from harm, rigour, research transparency, as well as preventing inaccurate findings from entering the evidence base. However, the social and environmental sciences and engineering fields, and safety outcomes therein, compared to healthcare are less regulated for several reasons: 1) the different risk of harm to humans 2) the diverse research methods based on precision and control within a given context and 3) variable governance systems across different countries, for example. The context of the field of study, application and industry cannot be overlooked in this respect.

Challenges to peer review standardisation in the safety space include:

- Safety research is conducted across different organisations, from academia to industry, NGOs, charities and government affiliates.
- While academic projects involving human participants, data and/or biological samples, are assessed university ethics committee for adherence to safety and legal standards, similar mechanisms are not guaranteed in other organisations.
- Furthermore, different countries have different legal frameworks for university ethical review, even in comparable income countries (*Goodyear-Smith, 2002*).

As no single body is responsible for safety review even at the country level, peer review should be purposefully built into project design and cycle by researchers. At Lloyd's Register Foundation initial checks on ethics, feasibility, suitability of the study team, study activities, financing and due diligence are conducted at the contract stage. However, beyond that, we have developed the Global Safety Evidence Centre advisory panel, from which smaller consultation groups are selected for any given project to provide input and provide documentation review. **Importantly**, the consultation group does not replace other external peer review mechanisms (see chapter 3 on ethical review) as the consultation group is involved from project initiation to completion, so is not fully independent.

Consultation groups are made up of varied professionals and practitioners with relevant method and/or topic expertise in the specific areas of the research. They have advisory capacity, review study documents and meet at 3 critical check points – before activities commence, the interim period and study completion.

In any discipline, research plan changes are common, and often necessary, due to operational issues or new data. Depending on the changes, research activities may be halted during this time to control for potential risks to participant safety, ethical considerations or data integrity. Changes to study plans should be reviewed at this time (see chapter 3).

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Pre-registration and transparent documentation

Due to researcher advocacy and journal requirements pre-registration is becoming more common; where the aim is to achieve research transparency, accountability and replicability. Pre-registration involves submitting study plans to an online repository before any study activities take place. Documenting study plans upfront helps control researcher degrees of freedom and bias. Preregistration is particularly important in fields that have replication issues – e.g. across the safety literature – where research approaches and analytical choices are broad.

In the medical field, pre-registration is essentially covered by the regulatory authorities as the operational team are accountable to any submitted documents. Without a specific regulatory body, pre-registration can be more important. In evidence synthesis research, pre-registration of the review protocol has become the standard – see *Evidence synthesis and review method guide*. Pre-registration is also strongly encouraged for primary evidence projects at the Global Safety Evidence Centre and for an RCT a pre-registered protocol and statistical analysis plan are considered mandatory. For other data collection projects, design and analysis documents are expected to be available at project initiation.

Study documents can be pre-registered privately or via open access platforms. Notably, ability to share company information will differ between organisations, with industry having more restrictions. However, privacy concerns can be mitigated by submitting redacted documents and open access registration is encouraged.

Below is a list of useful pre-registration repositories and sites for safety evidence research – it is not exhaustive, but resources are free:

- The best all-round preregistration repository for primary evidence is the Open Science Framework (OSF)
- For RCT studies in economics and social policy the American Economic Association (AEA) RCT registry can be used
- Wharton University has a range of open access pre-registration tools including their open access AsPredicted site (9-question pre-registration tool).

Other resources:

- Protocols.io is a protocol development, storage and sharing site. Useful for more technical protocol development but not pre-registration per se.

Primary evidence planning documentation

In primary evidence commonly pre-registered documents include:

Protocol:

- Project title, information and responsibilities
- Background and rationale including supporting evidence, why the study is important and what gaps are addressed
- Research aims objectives and research questions
- Research design
- Target population, sampling method, sample size and justification
- Data collection method, procedure, instruments, variables and metrics
- Metric and measures development and validation (if required)
- Pilot testing and purpose (if required)
- Ethical considerations
- Data analysis and storage (top level if supplemented with a statistical analysis plan)
- Data management (top level if supplemented with a data management plan)
- Timeline
- Limitations

Statistical Analysis Plan:

- Overview and responsibilities
- Analysis populations and groups
- Data handling, including missing data
- Statistical methods
- Interim analyses including timing and stopping rules (if required)
- Subgroup analyses
- Sample size justification
- Software
- Results presentation and reporting plans

Data Management Plan:

- Data description, data types and formats
- Data collection and documentation
- Storage and backup
- Data security and confidentiality
- Legal and ethical compliance – local and internation incl., GDPR
- Data quality assurance
- Data sharing and access
- Archiving and preservation
- Responsibilities

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Further reading

Goodyear-Smith, F., Lobb, B., Davies, G., Nachson, I., & Seelau, S. M. (2002). International variation in ethics committee requirements: Comparisons across five Westernised nations. BMC Medical Ethics, 3(1), Article 2. https://doi.org/10.1186/1472-6939-3-2

Kelly, J., Sadeghieh, T., & Adeli, K. (2014). Peer review in scientific publications: Benefits, critiques, and a survival guide. EJIFCC, 25(3), 227–243. https://pmc.ncbi.nlm.nih.gov/articles/PMC4975196/

GOV.UK. (2014). Clinical trials for medicines: Apply for authorisation in the UK. https://www.gov.uk/guidance/clinical-trials-for-medicines-apply-for-authorisation-in-the-uk

Lloyd's Register Foundation. (2025). Evidence review and synthesis methods guide. https://www.lrfoundation.org.uk/publications/evidence-review-and-synthesis-methods-guide

Center for Open Science. (n.d.). Preregistration. https://www.cos.io/initiatives/prereg

AEA RCT Registry. (2019). Social Science Registry. https://www.socialscienceregistry.org/

Wharton Credibility Lab. (n.d.). Credibility Lab preregistration. https://credlab.wharton.upenn.edu/

Protocols.io. (n.d.). Protocols.io. https://www.protocols.io/

The Turing Way Community. (2024). The Turing Way: A handbook for reproducible, ethical and collaborative data science. The Alan Turing Institute. https://book.the-turing-way.org/

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Chapter 3. Ethical principles and considerations for data collection

Authors: Konstantina Mirtzani and Emily Lambe

Why ethics matter in primary research

Primary safety research often involves people sharing experiences of accidents, hazards, unsafe conditions, organisational practices, or failures in systems intended to protect them. While generating evidence in these areas is essential to improving safety outcomes, poorly designed or poorly governed research can cause harm to participants, researchers, organisations, or communities, and can undermine trust in evidence altogether.

Ethical primary research seeks to protect participants, maintain scientific integrity, and ensure that evidence generation contributes positively to safety outcomes. Ethics are therefore not an administrative add-on, but a core component of good research practice throughout the project lifecycle: from design and data collection to analysis, reporting, and dissemination. This chapter sets out the key ethical principles and minimum expectations for primary safety research commissioned or supported by the Global Safety Evidence Centre.

Ethical principles for safety research

Ethical safety research is grounded in a small number of well-established principles that apply across methods, settings, and sectors. In practice, these principles must be interpreted proportionately and responsively, recognising that safety research is conducted in diverse social, economic, cultural, and organisational contexts.

Voluntary and informed participation

Why important: Protects participants from coercion, misunderstanding, or unintended pressure, particularly where power imbalances exist.

Safety example: Participants receive a clear participant information sheet and consent form written in plain language, are reminded they can withdraw at any time, and are reconsented if research activities change.

Maximising benefit and minimising harm ("do no harm")

Why important: Ensures researchers proactively prevent physical, emotional, social, and economic harm associated with participating in research.

Harm is not always obvious: it may be psychological (e.g., distress during discussions), reputational (e.g., being identified as a whistleblower), relational (e.g., tension with employers), or practical (e.g., lost wages to participate).

Safety example: Before data collection, the team completes a risk and harm assessment (including emotional, relational, and economic risks), sets stopping rules for participant safety and data integrity, and co-designs benefits with participants (e.g., accessible feedback, referrals, small safety improvements).

Fair opportunity to participate and benefit

Why important: Creates fair opportunities by considering barriers such as literacy, language, time, access, safety, or distrust. At the same time, inclusion must never override voluntary participation or expose people to risk.

Fair benefit means communities contributing data should not be merely "studied"; they should meaningfully benefit from the findings and engage in research. This could include sharing results in accessible formats, supporting capacity building, or informing interventions that improve safety outcomes.

This principle protects against extractive or exploitative research practices, especially relevant for workers who lack formal protections or operate in precarious economic or social positions.

Safety example: Participation is scheduled around shifts; transport/childcare are reimbursed; interpreters are used; opting to randomise by time in experimental studies so that all groups have an opportunity to potentially benefit; communities validate findings and receive results in accessible formats.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Confidentiality, privacy, and data protection[1]

Why important: Privacy and confidentiality protect participants' identities, experiences, and personal information. Ethical research requires safeguarding this information from misuse, unauthorised access, or unintended disclosure.

Confidentiality is particularly important in safety research, where individuals may report hazardous practices, unsafe conditions, or organisational failures. Mishandling their data could expose them to retaliation, job loss, or social consequences.

Participants should know:

- what data is collected
- why it's collected
- who will have access
- how long it will be stored
- how it will be anonymised or pseudonymised (depending on project and data- see UCL guidance)
- how and where results may be shared

Protecting privacy is both a legal requirement and a moral obligation grounded in respect and trust.

Safety example: Anonymisation goes beyond names, removes or re-codes job role, unit, shift, and incident details in small workplaces; data is encrypted, access is tiered, retention and destruction are defined; data use is explained in clear terms

Scientific integrity[2]

Why important: Ethical research is scientifically sound. Weak, poorly designed, or biased methods can produce misleading, harmful, or unusable findings.

Safety example: Methods are predefined and proportionate to risk (e.g., anonymised surveys for sensitive topics); deviations are documented; reflexivity logs capture assumptions; triangulation supports findings; limitations are disclosed in all outputs.

1. (Detailed data protection guidance will be in a separate section; this section simply establishes the principle.)

2. See chapters above

Social and environmental responsibility – research

Why important: Research activities should avoid causing environmental damage or contributing to unsafe or unethical local practices. Where possible, benefits are maintained after study completion.

Safety example: Local partners are trained and paid fairly; travel is minimised; data collection avoids disrupting livelihoods; green procurement is used; an exit strategy prevents extractive engagement and supports sustained local capacity.

Ethics review and oversight recommendations

All data collection projects with human participants should be subject to ethical review; before data collection and on an ongoing basis. Ethical principles also apply to data analysis projects in terms of data integrity, intellectual property, environmental impact, and societal consequences.

Generally, to ensure ethical standards are met and maintained documentation at two levels should exist:

1. Context relevant processes and policies and the organisational level
2. Project-specific independent ethical review

A core function of pre-study independent peer review is ethical review. Safety research is performed by many different types of organisations (e.g. NGOs, academia, industry) in different countries, and with that different processes will apply (as outline in chapter 2). University-affiliated projects, for example, are subjected to the university ethics committee, and often it is most effective to partner with an institution that has a formal ethical review process. However, for projects not affiliated with a university or relevant ethics committee, ethical review and consultation can be sought from independent bodies (varied fees). Examples of such bodies across the globe are:

- UK – IREC - Independent Ethics Research Committee
- UK – SRA – Social Research Association
- Europe - EUREC - European Network of Research Ethics Committees. Umbrella network connecting independent and institutional RECs across Europe
- Africa – AIEC - Africa Independent Ethics Committee. Research network for social change across Africa.
- New Zealand – IHREC - Independent Human Research Ethics Collective

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Several smaller country level organisations and services can be found through online searches and LinkedIn.

At Lloyd's Register Foundation, preliminary checks are performed at the application stage, however it is the responsibility of the managing organisation to ensure that the aforementioned mechanisms are adhered to and relevant documentation available.

Any changes to data collection, storage or research plans should be subject to ethical consideration and updated informed consent if necessary.

Ethical considerations in experimental research and trials

Any research that has an intervention element should be subject to increased ethical inspection. For ethical considerations relating to experimental research and trials in the safety space see chapter 7.

Further reading

Buchanan, D., & Warwick, I. (2021). First do no harm: Using 'ethical triage' to minimise causing harm when undertaking educational research among vulnerable participants. Journal of Further and Higher Education, 45(8), 1090–1103. https://doi.org/10.1080/0309877X.2021.1890702

Adroit Associates. (2024, February 8). Safeguarding and do-no-harm principles for vulnerable populations in conflict-affected settings. https://adroitassociates.org/publications/2024/safeguarding-and-do-no-harm-principles-for-vulnerable-populations-in-conflict-affected-settings

Jawa, N. A., Boyd, J. G., Maslove, D. M., Scott, S. H., & Silver, S. A. (2023). Informed consent practices in clinical research: Present and future. Postgraduate Medical Journal, 99(1175), 1033–1042. https://doi.org/10.1093/postmj/qgad039

Klykken, F. H. (2022). Implementing continuous consent in qualitative research. Qualitative Research, 22(5), 795–810. https://doi.org/10.1177/14687941211014366

University College London. (2025). Anonymisation and pseudonymisation of personal data. https://www.ucl.ac.uk/data-protection/guidance-staff-students-and-researchers/practical-data-protection-guidance-notices/anonymisation-and

Kretser, A., Murphy, D., Bertuzzi, S., Abraham, T., Allison, D. B., Boor, K. J., Dwyer, J., Grantham, A., Harris, L. J., Hollander, R., Jacobs-Young, C., Rovito, S., Vafiadis, D., Woteki, C., Wyndham, J., & Yada, R. (2019). Scientific integrity principles and best practices: Recommendations from a scientific integrity consortium. Science and Engineering Ethics, 25(2), 327–355. https://doi.org/10.1007/s11948-019-00094-3

Resnik, D. B. (2025). Privacy and confidentiality. In The ethics of research with human subjects: Protecting people, advancing science, promoting trust (2nd ed., pp. 199–219). Springer. https://doi.org/10.1007/978-3-031-82757-0_6

Restubog, S. L. D., Schilpzand, P., Lyons, B., Midel Deen, C., & He, Y. (2023). The vulnerable workforce: A call for research. Journal of Management, 49(7), 2199–2207. https://doi.org/10.1177/01492063231177446

University of Michigan. (2018). Informed consent guidelines & templates. https://hrpp.umich.edu/irb-health-sciences-and-behavioral-sciences-hsbs/informed-consent-guidelines-templates/

Digital.gov. (2025). Plain language guide series. https://digital.gov/guides/plain-language

UK Research and Innovation. (2023). Research ethics guidance. https://www.ukri.org/councils/esrc/guidance-for-applicants/research-ethics-guidance/

Social Research Association. (2021). Research ethics guidance. https://the-sra.org.uk/SRA/SRA/Ethics/Research-Ethics-Guidance.aspx

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Chapter 4. Data collection tools, validation and data types

Author: Emily Lambe

This short chapter explores data collection methods and validation processes in relation to different data types. Collecting data through validated methods helps ensure more good data not just more data.

Tools for safety primary evidence research

Questionnaires and surveys are essential tools for primary data collection and can have both quantitative and qualitative components. Questionnaires tend to be a more specific data collection tool than a survey which is a broader analytic tool, however similar principles apply to both so may for this purpose be used interchangeably.

Questionnaire development should be a systematic process to ensure that the end-product consistently measures (is reliable) what it planned to (is valid). For this reason, using tested and known surveys, with shown external and internal validity, is always preferable and far quicker for project design and initiation.

Validation of surveys and questionnaires

For standardisation of metrics and measures, it is essential that the research tools used are validated. It first worth checking for existing suitable and validated surveys. But, there are times when development of a topic and context specific questionnaire, particularly to expand or explore new research topics in safety, is needed. New surveys and questionnaires should go through a development and validation processes, i.e., piloting, statistical assessment (quantitative items) and peer review, to ensure reliability and validity in the following ways:

- Face validity – does the survey seem appropriate at face value? Length, format, topics considered.
- Content validity – do these questions cover all aspects of the topic or domain?
- Construct validity – does any question or set of questions accurately measure the construct (e.g. sleep disturbance) intended?
- Criterion validity – do the measures correlate with an existing well-validated tool? Does it collect similar data, and can it predict similar outcomes?
- Internal consistency – how similar are survey items as a unit?
- Repeatability/ test-retest – does the survey produce the same data if completed at a second time point by the same people?

Self-report versus observer-rated data

In safety research, often those completing the surveys are observers of the workers or practitioners, rather than the workers or practitioners themselves, which could be for a variety of reasons (e.g., different management systems, worker literacy, or simply because the measure is designed for audit and inspection purposes). Where the survey is completed by one or more observers (raters) of the behaviour or system, the following should also be considered:

- Intra-rater reliability – does the observer produce the same score at another timepoint? A similar but not identical concept to test-retest reliability.
- Inter-rater reliability – do two or more observers produce results that align?

Rater reliability can also be applied to non-human observations. For example, manually rating number of weather events, where subjective bias can play a role.

Each survey item needs to be considered for its value. The table below includes common examples of how reliability and validity can be assessed. **However, it is paramount to note that the statistical methods below cannot be applied to all surveys. Survey research methods and statistics is large field, and researchers should find the best validation methods for the type of data they are planning to collect.**

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Table 3. Validation principles for new research tools

Construct	Test	Description
Face validity	Peer review	Top level checks via colleagues.
Content validity	Peer review/ content validity ratio	Expert review. CVR is simple metric of based on how many experts view a survey item as essential.
Construct validity	Factor analysis	Tests whether items load on to a higher latent factor
Criterion validity	Pearson's r test (for continuous data, linear effects) or other correlation coefficient (non -linear relationships)	Tests how much data varies with other data. Use should be based on satisfied assumptions.
Internal consistency	Cronbach's alpha (for unidimensional surveys)	Tests how similar survey items are. For multidimensional surveys other methods should be used.
Repeatability/ test-retest reliability	Intra-class correlation (continuous/ ordinal data)	Practice effects and therefore timing is an important consideration.
Intra-rater reliability	Intra-class correlation (continuous/ ordinal data)	Practice effects and therefore timing is an important consideration.
Inter-rater reliability	Intra-class correlation (continuous/ ordinal data)	All raters should have the same information/ guides.

Various books and guides on survey research methods exist but some accessible resources and examples are:

>> Guide to questionnaire development and validation – this paper covers several steps and methods.

>> UK government survey validation targets – Recently published UK government guidance on survey development using the agile project management approach.

>> World Risk Poll Methodology – example questionnaire development methodology in the Lloyd's Register Foundation World Risk Poll

Qualitative interviews and focus groups are also key tools in safety research for understanding relevant, context dependent and practitioner-led perspectives. Qualitative methods are not a focus of this guide, however for completeness the general principles and requirements are outlined here.

Interviews provide rich and person-centred data and have a variety of uses in the safety space. For example, in development and evaluation of safety interventions, or as scoping practices to understand risk, need and attitudes towards an activity or technology, i.e., exploratory research.

The questioning approach will differ depending on the type of data sought but the common approaches are:

Structured interview

Structured interviews bridge qualitative and quantitative approaches in so far as interview questions are standardised, ordered and often scored. For example, an opportunistic in-person survey on road safety.

Semi-structured

Semi-structured interviews involve open ended question, that can be adapted or added to during the interview and are useful when questions/interests are pre-set, but depth and flexibility are required for data richness.

Unstructured

Unstructured interviews operate more like a conversation. The interview might start with an opener such as 'can you tell me something about the culture at work?' but generally the interview is led by the interviewee.

Focus groups

Focus groups involve a small group of people (6-10 usually) who come together to discuss a specific topic and provide subjective experience or option, with a moderator to guide a session and keep discussion on topic.

Roundtables

Not a traditional research method per se, but popular in the safety space. Roundtables are similar to focus groups but involve a group of experts with the aim of finding solutions or sharing insights and have a less defined analysis framework. For example, the Global Safety Evidence Centre was interested in the *current and future status of safe work* and held a roundtable event to explore this topic.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Validation of interviews and qualitative data

The number of interviews performed will vary depending on the study requirements and population heterogeneity. However, in choosing the number interviews:

- The principle of data saturation is important to consider, i.e., the point where no further themes or insights can be extracted from more information.
- Depth (interview length and questions) versus breadth (how many different interviewees) is another important consideration.

Validated interview research plans are also mandatory and concerned with trustworthiness rather than quantitative reliability and validity. Methods for qualitative validation include piloting, triangulation of data sources and checking-back on results with participants.

There are, of course, other ways to collect data than surveys and interviews, however a key point of this chapter is that whichever tools are used to measure data of interest, those tools and methods should be assessed and validated. In the engineering view, instrument calibration, traceability, measurement uncertainty, sensor accuracy and drift, inter-operator and inter-lab repeatability, and validating models against physical data, would be considered.

Data types

It is essential to understand the type data that will be collected as this allows the correct analysis techniques to be planned. Some tests are more robust to errors than others, however, choosing the wrong test could result in incorrect test statistics and invalid conclusions.

Qualitative - non numerical data

- Free text – usually included for qualitative elements of a questionnaire or survey
- Recordings – interviews are recorded to allow later transcription
- Other non-linguistics and non-numerical information, such as photographs and drawings, but this is less common in the safety space.

Quantitative - numerical data

- Discrete data – consists of numerical whole values. For example, a person can only sensibly be counted as 1, not 0.5.
- Categorical data:
  - Nominal – categorical coding and the simplest form of measurement, where data categories have no quantitative value, relationship to each other or ranking, e.g. 'manufacturing', 'fishing', 'technology'. Numerical 'dummy' codes are often used to represent categories but are not meaningful values in themselves.
  - Ordinal – categorical data that does have ranking/order but the difference between values is unknown. For example, 'low', 'medium', 'high'. A common mistake is to confuse ordinal data for interval data. However, these data types have distinct assumptions.
- Continuous data - can be any value within a range. It can be an infinite number of possible values.
- Continuous data:
  - Interval – ordered data that has equivalent intervals between values. For example, the difference between 1 and 2 is the same as the difference between 1245 and 1246.
  - Ratio – interval data that has an absolute zero value, i.e. a value of 0 means absence, which allows meaningful ratio conclusions. Age, for example, where a 20-year-old is twice as old as a 10-year-old.

A good understanding of different data types allows accurate analysis, which is covered further, for quantitative data, in the next chapter.

Further reading

Kishore, K., Jaswal, V., Kulkarni, V., & De, D. (2021). Practical guidelines to develop and evaluate a questionnaire. Indian Dermatology Online Journal, 12(2), 266–275. https://doi.org/10.4103/idoj.idoj_674_20

Government Analysis Function. (2023). Survey development toolkit: An overview of the research objectives and methods for each Agile phase of survey development. https://analysisfunction.civilservice.gov.uk/policy-store/survey-development-toolkit/

Lloyd's Register Foundation. (2024). World Risk Poll 2024 report methodology. https://www.lrfoundation.org.uk/sites/default/files/2024-06/World%20Risk%20Poll%202024%20Report%20Methodology.pdf

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Chapter 5. Quantitative data characteristics, distributions, probability and statistics

Author: Emily Lambe

Different data types have different characteristics and distributions, which is why consideration is important when it comes to hypothesis testing and analytical choices. Different types of data should be analysed with tests that model different probability distributions and assumptions; principles and examples of which will be explored in this short chapter.

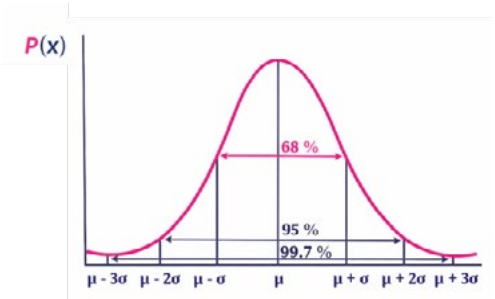
- Hypothesis – a specific prediction based on the direct of the relationship/effect that is expected.

Distributions

Theoretical and practical (experimental) probability and probability distributions have been debated for centuries. What follows is a brief introduction and reflection on the link between data types and probability theory. At a simplistic level, two distributions, normal and binomial distributions, are common for analysing continuous categorical data respectively.

Normal distribution (Gaussian)

The normal distribution is represented as a bell-shaped curve, where the mean, median and mode of dataset or variable are the same (perfect gaussian), and the highest or densest concentration of values is in the centre of the curve, around the mean.

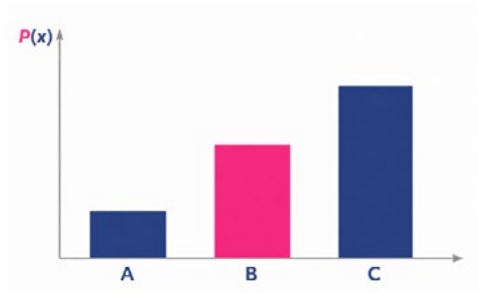


Where:

- P(x) is probability function
- μ is mean
- σ is standard deviation (spread from the mean)

Binomial distribution

The binomial distribution is appropriate when data falls into distinct buckets, for example pass or fail, yes or no. Distribution beyond two categorical levels is generally referred to multinomial distribution.



Where:

- P(x) is probability function
- x axis are categorical buckets

Other distributions, uniform and Poisson, are used less frequently in basic statistical analyses, though the Poisson distribution is useful for modelling number of event within a certain interval, for example, accidents within a mile or earthquakes within a year. This blog from University of Sheffield provides a practical summary *>> Probability Distributions*.

Data is also characterised as parametric or non-parametric. Parametric analysis requires data to follow a specific distribution (usually normal) whereas non-parametric does not.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Probability – Frequentist and Bayesian methods

Broadly, probability is the likelihood of an event or effect occurring (or not) and has a value between 0 (impossible) and 1 (certain). In primary research and data analysis it is important to be able to understand the level of uncertainty in a given finding (and therefore the conclusions that can be made about that finding). Probability is a mathematical tool for this and for predicting outcomes that may or may not occur. In the traditional frequentist perspective probability is seen as the proportion of times an event, effect or observation occurs given an infinite (or very high) number of attempts.

Traditional (frequentist) probability views chance as static. For example, if you were to roll a die 6 times the chance of getting any number 1-6 would be equal each time. In contrast Bayesian probability finds that the chance of an event is related to events that have happened before (prior events). While traditional statistical significance relies on the null (i.e. assumption of no relationship or effect), Bayesian statistics are concerned with the truth of a hypothesis given a dataset.
>> Bayesian Statistics (IBM) – a mid-level explanation.

Use of frequentist probability and statistics is generally the standard and is included in statistical software as such. However, many programming (e.g., R-Studio, Python) and Graphical User Interface (GUI) based (e.g., Stata, Jasp) products now include Bayesian functionality. Where confidently understood Bayesian methods can be a good option for data analysis.

Statistical analysis – an introduction

Advanced methods and statistics are beyond the scope of this guide. However, the aim is to share some key considerations and approaches for different audiences in the safety space.

Statistical analysis allows testing of whether the patterns or 'effects' in the data are likely to reflect real world phenomena, as well as whether the data we have allows us to predict other data and phenomena. It can be used for hypothesis testing or exploratory analysis, however as mentioned in chapter 2, it is essential to set that intention apriori and in any study plans. All statistical tests should be run in line with the assumptions of the data and test, which is why it is essential to correctly classify data and variables type ahead of time.

The most common ways to do this are:

- First step to understanding data is to visualise it, via histogram, boxplot, scatterplot for example.
- Second step is to generate descriptive statistics e.g. mean and median and standard deviation, and check the features of the distribution they follow.
- Third step is to decide on the probable statistical tests and check the assumption of the test are met, e.g. normality, skew, kurtosis and variance. Shapiro-Wilk test and Levene's test are commonly used.

Regression is a strong tool for statistical analysis, used across different fields, and should be employed in the safety space, where appropriate. Its strength is that different data sources and types can be incorporated. Regression analysis involves minimising error in the regression model to find the 'best fit' of the data. Regression models for different data types include:

Table 4. Shows examples of different regression models for different data types

Model	Data type
Linear Regression	Continuous linear
Logistic Regression	Categorical
Non-linear Regression	Continuous non-linear

Significance

Statistical significance refers to the probability observing an effect in the data, based on the null hypothesis (usually that no, or the opposite, effect exists in reality) being true, i.e., observing an effect when there is not one, otherwise called a false positive. The level which is selected as appropriate for the significance of a test result is quite arbitrary but there are norms by field. For example:

- Social sciences and psychology $p<.05$ – $p<.01$
- Medicine and clinical trials $p<.001$
- Engineering and safety science - no status quo due to diverse methods. Significance level, where one is used, should be justified for the study design, context, multiple testing and the consequences of false-positive and false-negative conclusions (see below).

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

A p-value of.05 means that chance of observing an effect in the data when there is not one is 5%. A highly significant result (i.e., small p-value) generally gives some confidence that the effect observed is a true one (i.e., not a false positive), but significance is not the only factor that should be considered (see chapter 2 on bias and systematic effects that should be planned for in research design).

Historically, much emphasis has been put on statistical significance in the literature and evidence base. However, it is important to note that significant does not mean true nor does it mean good. It is recommended that p-values are reported with other statistics reflecting variance, effect size, sample size, goodness-of-fit and range. The office for National Statistics has useful online pages on supporting statistics: *>> Uncertainty statistics*

Multiple comparisons

The chance of significant results being due to false positives increases with the number of tests run and results generated; known as the multiple comparisons problem. A common and recommended correction methods for this in regression analysis is the Bonferroni correction method (Bonferroni, 1936). However, the choice of correction method will relate to the data and tests performed. In experimental research, the number of tests run should be planned and generally constrained to the amount required to answer a given question.

Confidence intervals

Confidence intervals indicate a range of plausible values for a given statistic or statistical parameter e.g., the mean or beta coefficients within a regression analysis. In this respect confidence intervals reflect the amount of uncertainty (margin of error) we have in a given result, where wider confidence intervals reflect a less certain estimate. Use of confidence intervals is recommended in general. In the safety space, particularly for research areas where data (and statistical estimates) are likely to be highly variable, human factors/self-report research (in contrast to mechanical engineering) for example, reporting confidence intervals is recommended.

Power

Statistical power refers to the probability of not observing an effect in the data when there is not one, i.e., correctly rejecting the null hypothesis. Power also relates to sensitivity, i.e., sensitivity to detect and effect in the data when true in reality. The standard is for power to be set at.8 i.e., 80% but again consideration of the research question, data and standards within the field should be taken into account.

Consideration of power and significance is essential to quantitative project design and should be included in the research proposal and plans. Different formulas exist for calculating a required sample size, depending on the anticipated size of the sample e.g. +/- ~ 100. Knowledge on the population, accessibility and previous literature should be used for this. More information can be found in this guidance:

>> Guidance on selecting a sampling technique and sample size estimator

Comprehensive and freely accessible Graphical User Interface (GUI) software to calculate sample size for a range of tests is available with supporting guidance:

*>> G*Power, supporting information and guidance (Mac and Windows)*

Further reading

University of Sheffield. (2021). Probability distributions. https://dataviz.shef.ac.uk/docs/18/03/2021/LearningPath-Statistical-Modeling-1

Noble, J. (2025). Bayesian statistics. IBM. https://www.ibm.com/think/topics/bayesian-statistics

Office for National Statistics. (2021). Uncertainty and how we measure it for our surveys. https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/uncertaintyandhowwemeasureit

Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R. Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 8, 3–62.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Chapter 6. Case studies, observational and quasi-experimental research

Author: Emily Lambe

This chapter attempts to unpick the important distinction between causal (experimental) and non-causal (case studies, observational and quasi-experimental) research frameworks, particularly relating to human organisational and social contexts. It should be acknowledged that in disciplines where precision and control are extremely high, such as engineering, mechanistic, observational and quasi-experimental testing can support causal conclusions.

Statistical analysis techniques can be versatile and data analysis must be performed in relation to the research design and method of data collection. For example, a regression analysis can be applied to a case study, observational survey data or experimental data, however the conclusions that can be made from the results are quite different. Therefore, statistical analysis should be considered in the context of 1) is it exploratory or hypothesis testing? 2) what framework has the data been collected in – observational or causal? See chapter 2 for context.

Case studies

Case reports/studies are, usually exploratory, primary evidence methods that can be qualitative and/or quantitative and provide useful evidence when data is difficult to access, collect and/or when very little evidence already exists in a given area. Case studies are not subject to the same requirements as more rigorous primary evidence methods, a justified sample size or control group in quantitative method for example, but are an important starting point in research discovery and often provide different insights to large scale operationally challenging studies. A guide to effective case studies can be found here: *>> Guide to effective case studies*.

Participatory Action Research (PAR)

PAR is a research approach that operates on a fundamentally different premise to researcher driven quantitative experimental research methods. The core idea is that the participants are also researchers so that rich, context specific information is collected and built without the enforced direction, bias or preconceptions of an external researcher. This approach strongly aligns with the Global Safety Evidence Centre's stakeholder research capacity, as well as with Lloyd's Register Foundation's locally led, global strategy; local voices provide the most relevant and useful information for the context. Data collection tends to be more subjective, which aligns with qualitative methods (covered in other/future guides). However well-considered research plans and the allocation of responsibilities are still required upfront for these studies. For further information on PAR: *>> University of Reading PAR toolkit*

Observational research

Observational research involves exploring or testing data that has not been collected in a causal framework, i.e. without manipulation or control of variables or conditions, but where sufficiency powered dataset can still provide strong insights. The most common quantitative observational tool is the survey. Large surveys can be very powerful for gaining insight into a broad range of phenomena. For example, the Lloyd's Register Foundation *World Risk Poll*, that has survey data from 147,000 people in 142 countries.

Observational research is highly useful for identifying and exploring patterns, trends and non-causal relationships between variables. Types of observational study designs include:

- Cohort studies – observing a particular group or groups over time (prospective and retrospective)
- Case-control studies – passively comparing two or more groups (retrospective)
- Cross-sectional – analysing data from a population or subset at a certain timepoint

In observational research the control group is identified rather than created, and can be based on:

- Predefined groups (i.e. from previous studies/cohorts),
- Randomly selected subset of the data,
- A subset of the data matched by certain characteristics, e.g., age, income, geographical location.

Due to the uncontrolled nature of observational studies, more data is required (than in experimental deigns) for statistical analysis to be appropriate and powered. If underpowered, it is best to explore observational dataset with descriptive statistics and visual analysis only.

Large observational data analysis can require considerable data preparation and management. This is, again, due to uncontrolled nature of the design and it is recommended that the data is well understood prior to any analysis. It is important to weight the data and/or include any potentially contributing (3rd party) variables in the analysis with the aim of controlling for or 'partialling-out' the influence of these variables on the relationship of interest.

A key advantage of observational studies is, because no manipulation is involved, data are easier to obtain, making larger datasets possible. However, it is vital to keep both researcher bias and the premise that data and variables moving together are not necessarily linked, in mind. Correlation ≠ causation.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Causality in research

Ability to make causal statements requires a design that includes the following elements:

- A manipulation – a change in input (sometimes referred to as intervention – see below)
- Temporal importance – cause (the manipulation) must precede effect (measured outcome)
- Purposefully collected data – data must be collected in accordance with the study design
- Control – other potentially contributing factors to a change in outcome must be accounted for in design and analysis.

In social research and policy, the term 'intervention' is used more frequently than manipulation. Technically speaking a manipulation is a highly controlled change in an independent variable and is language often used in experimental research and the hard sciences e.g. physics, engineering. Broadly, an intervention refers to a change in something with the anticipation of an improvement in certain outcomes, but is often used to mean:

- an experimental manipulation
- a series of activities
- a programme of change underpinned by a theory of change (ToC)

While these definitions are used interchangeably, there are different connotations for each in terms of scale, research methodology and conclusions that can be made and it is important to be clear on the justification at the design stage of a project. For impactful research, it is recommended that all research activities are considered in relation to a theory of change. The Global Safety Evidence Centre's current ToC can be seen on page 10 here.

Quasi-experimental studies

Quasi-experimental design covers a wide bracket of studies that are not causal but can be indicative of causal relationships. The reason is that they involve a manipulation or intervention and measure of change to the outcome but omit random assignment so it is not possible to control certain confounders. Quasi-experimental approaches provide vital evidence where trials are unfeasible or not yet justified.

In quasi-experimental and experimental research, a control group is purposefully designed or created. Commonly, there are self-control groups, i.e., in repeated measures designs where a person or group experiences both intervention and control conditions, or a totally separate control group. There are advantages and limitations to both, mostly around order effects and variability (see chapter 7), but the method will depend on what the research questions are.

Types of quasi-experimental designs include:

- Non-equivalent groups (independent control group). Assumes non-matched, groups though some matching could be performed.
- Interrupted time series (self-control group)

There are ways to consider both order and group effects in the design, for example using a cross-over design. However, as considerable overlap between quasi- and experimental designs exists this is explored in chapter 7.

Trap of assumed causality in organisational and social research

A common issue with quasi-experimental designs is the trap of causality. As a manipulation/intervention is involved it is tempting to draw conclusion statements like 'the intervention improved safety outcomes'. However, it important to not assume causality and ensure any conclusions made should be in the context of potential miscellaneous contributing factors.

Studies are often considered experimental studies, or controlled trials, if there is considerable control over group allocation, the manipulation and possible interacting/confounding influences/variables. However, a true experimental design in social and organisational contexts involves randomisation of the allocation to different groups which are explored in the next chapter.

Further reading

What Works Wellbeing. (2024). Guide to effective case studies. https://whatworkswellbeing.org/resources/guide-to-effective-case-studies/

University of Reading. (2018). Participatory action research. Participatory Methods. https://www.participatorymethods.org/methodology/participatory-action-research/

Lloyd's Register Foundation. (2025). World Risk Poll. https://www.lrfoundation.org.uk/wrp

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Chapter 7. Experimental research and trials

Author: Emily Lambe

Foreword: need for experimental research in the safety space

Experimental research describes a scientific and systematic approach to research enquiry which involves a change in a variable, set of variables or activities (see chapter 6), to assess the effect on a relevant outcome of interest, or range of. Experimental research is prospective rather than retrospective like observational research.

Randomised Controlled Trials are type of experimental research where increased experimental control is gained through a randomisation process; common place in the medical field and a foundation of evidence-based healthcare. Trials are essential to the medical industry as it is critical to know whether a drug 'works', or not, and importantly whether there are negative effects from the drug and the likelihood of those occurring.

Application is not limited to medicine and over the last 50 years trials have been increasingly implemented in psychology and behavioural science fields, economics and social policy, and business and technology.

The safety space includes broad and complex social and technical systems, with few one-size fits all solutions. This means most safety guidance is based on experience and expert opinion, and for good reason too. However, trials are particularly useful for evaluating the outcomes of social, digital, environmental and workplace interventions.

Trials are not the only route to causal conclusions across the safety space. As mentioned in previous chapters, in disciplines where control is high such as engineering systems, casual evidence can be generated through quasi-experimental methods. But in the social context where variability is high, trials are important.

The Global Safety Evidence Centre's 'what works' pipeline started with a review of OSH interventions for safety (a review of reviews), which reflected the known heterogeneous nature of the OSH evidence base, OSH outcome measures and lack of rigorous intervention studies. By building the experimental evidence base, the aim is to generate an accessible toolkit of evidence-based activities and interventions that are shown to work for improving safety outcomes, which can be achieved through experimental research and randomised controlled trials, together with evidence synthesis.

Safety factors and contributors are not static nor necessarily unidirectional but will rely on cause and effect. RCTs can complement systems thinking by helping unpick which activities within a given system are effective. RCTs also support more robust statistics by better satisfying the assumptions of statistical tests and probability theory. For these reasons, it is hoped that this chapter encourages more RCTs in the safety space if appropriate given the research questions and current evidence base (see chapter 2).

Effectiveness is a cornerstone of experimental research and showing what interventions work. In research, the term effectiveness has a specific meaning, relating to data collected in a causal framework where differences in outcomes are shown.

Like other sections of this guide, this chapter will provide principles, concepts and design options for trials in safety. Pivotal texts if you are considering a trial are:

>> Effectiveness and Efficiency - Random reflections on healthcare services

>> The Ochre Book: Designing and delivering randomised trials for social policy

Units of study (analysis)

Due to the common use of trial design in the medical field, as well as psychology and social research, where those being studied are people or patients, terminology in the literature used to explain trial design is often based on participants, particularly at the individual level. However, units of study will not always be participants, and trial design can be applied to many contexts.

In engineering research, the units of study could relate to objects, materials or digital entities; one application would be testing a new technology in different digital or physical stress conditions, for example. In the environmental safety field, trial design could be used to investigate different sustainability programmes on environmental safety factors, like chemical contamination, for example. Therefore, in this chapter, you will observe 'participants', 'individuals' and 'units' terms being used interchangeably. However, it should of course be noted that there are different ethical considerations for human and non-human RCTs (see chapter 3).

Randomisation Theory

Randomisation is a fundamental method for control of systematic differences and certain biases. The idea is that each person or unit of study has an equal probability of assignment to any given condition (experimental or control) and therefore given an appropriate sample size other characteristics or confounders will be equally spread between groups.

Many strategies for randomisation exist. Some are:

- **Simple randomisation** – assigns individuals to groups by a 'flip of coin' approach, which is easy to implement but can produce unequal group sizes and needs a large sample.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

- **Blocked randomisation** – assigns individuals to blocks of equal sizing ensuring groups are balanced.
- **Stratified randomisation** – the population is first grouped into strata, which can be useful for subgroup analysis.
- **Factorial randomisation** – allows randomisation more than once, to series of conditions, without diminishing power.
- **Cluster randomisation** – randomisation at the group level rather than individual level but requires consideration of cluster size.
- **Adaptive/minimisation randomisation** – advanced 'along the way' randomisation methods where allocating participants to the more effective, successful or ethically beneficial arm, or based on the covariates of existing participants, in order to maintain balance or stratification, is desired.
- **Time-based randomisation (waitlist randomisation)** – all groups experience the intervention as the timing of the intervention is randomised, which can be ethically optimal but requires additional resources.

Advanced software exists to help with randomisation *>> Randomisation software - York University*

Blinding

Randomisation protects against certain biases, such as allocation bias, when paired with a blinding procedure, i.e. where participants are unaware of the conditions they are assigned to. Double blind studies, where both participants and researchers are blinded to group allocations, are even better for allocation bias protection.

In clinical trials the double blind is an essential part of experimental control in drug-placebo studies. The double blind is based on the placebo and treatment being indistinguishable i.e, the placebo (no drug) appears the same as the drug. Use of a placebo condition, ethically, must be based on there being no established effective intervention or standard already, minimal risk, close monitoring and possible early termination.

In the safety space 'placebo' trials have limited utility because the treatment and control groups are unlikely to be indistinguishable and therefore the double-blind not feasible. For example, it would be obvious to a researcher, employer or manager if they were providing training with a risk assessment component or training without that component.

Selecting an appropriate trial type depends on the research questions, hypotheses and population studied. The next section provides a high-level overview of different trial types that might be used. Researchers often utilise a combination of approaches.

Trial types and principles

Parallel group trials refer to control group comparison studies more broadly and the terminology applies better to the safety space. Regardless of terminology, efforts to keep which condition participants are in from them and each other is still worthwhile due to spill-over effects (see below). Third party administrators can be useful in these circumstances.

Superiority trials are interested in whether an activity or programme is better i.e., more effective, than the control. Placebo trials are a type of superiority trial where the aim is to establish whether the intervention is better than no intervention. Though safety standards differ by geography, industry and organisation, most workplaces will have some safety practices in place, making the concept of no intervention often unrealistic and potentially unethical.

The other type of superiority trial that is likely to be more suited to the safety space is where the control group is the current standard practice, which can be better ethically and more practical but requires consideration of possible baseline and interaction effects.

Equivalence trials and **non-inferiority trials** are interested in whether an activity or programme is as good as the control, i.e. the standard or current practice. One use, for example, would be testing the effectiveness of a new lower cost safety activity, where superiority might not be expected or required, with the current standard. This approach can support the mobilisation of resources that could then be used on other safety activities or elsewhere.

Superiority, equivalence and non-inferiority trials have different statistical assumptions based on the framing of the null-hypothesis and expected direction of effect. These have implications for power and sample size requirements and should be considered in design.

Spill-over and Leakage

In social research, design of the treatment and control groups requires careful consideration due to possible spill-over and leakage effects. Spill-over effects are when elements of the intervention spill-over into the control group. For example, giving training on safety culture to one or more groups, might improve safety culture measures across the whole company, including the control group, due to a trickle effect of behaviours from the treatment group.

Leakage effects occur where running the trial disproportionately disadvantages the control group, or other, so that the benefits to the treatment group are offset. For example, an initiative to improve safety outcomes for formal seafarers could be offset if high-risk activities are redistributed to informal workers.

Risk of both spillage and leakage effects should be considered in trial design.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Missing data and attrition

Missing data and attrition (drop out before study completion) are threats to statistical power. Though there are techniques to address missing data (see *the Turing Way*), with varied success depending on the reason for missingness, potential data loss, drop out and consequences of that should be considered in the design of the trial.

Baseline measures

Baseline (pre-test) measures are an important consideration in trials, particularly if you are interested in change from baseline and within-subjects factors (see below). Baseline measures can also be used for statistical modelling, i.e. as covariates, to control for differences between groups. It is also often useful to know what the starting point was, both in relation to understanding equivalence of groups and for long- and short-term monitoring purposes. Measures and metrics should be described in the trial protocol, i.e., selected upfront, and have shown validity and reliability in measuring what is intended (see chapter 2).

Between and within subjects/groups designs

A between-subject (or between-groups) design is where the unit of study, participant or group only experiences one condition in the study, either a treatment or a control group. Whereas, a within-subject (or within-groups), or repeated measures design is where the unit of study, participant or group participate in more than one condition (often all conditions).

Within-groups designs minimise random error comparatively but might not always be feasible. There is also the possibility of order effects, where the effects of being in a given condition carry over to another condition. Hence, timing is a large consideration in within-subjects designs.

Pre-test post-test and post-test only designs

Baseline measures are fundamental to pre-test post-test designs which allow within-groups changes, and generally changes over time, to be assessed. The best way to implement pre-test post-test is before and after each treatment or control period, not just at the beginning and end of the study, given that there could be differences over time.

There are some instances where pre-test or baseline measures are not possible and post-test only design is the better option. For example, if collecting baseline measures would be unethical, unfeasible or introduce detrimental testing effects.

If the design includes strict random assignment and a large control group, not having a baseline measure can be less important for covariate control, but is still important for measuring within-subject/group change, should that be a study priority.

Factorial trials

It is possible to implement a pre-test post-test design and control for the effects of the pre-test, i.e. testing effects, using Solomon Four group design. This design requires at least four groups and allocation to one of the levels in each condition, i.e. post-test only or pre-test post-test, and intervention or no intervention.

The Solomon Four group design is a version of a factorial trial design. Factorial trials allow the investigation of more than one safety activity, intervention, or a programme of activities in a single trial. Though factorial trials can be complex to interpret they allow the investigation of combinations of interventions as well as any given activity on its own.

Cross over trials

Cross over trials are a repeated measures design where participants, or units, are allocated to a sequence of treatments and control periods in a randomised order. In this way each person or unit acts as their own control and order effects are accounted for by the randomisation of the sequence. In a cross over design, you would expect pre-test and post-test measures to be collected before and after each treatment period.

Technology trials

- Decentralised or digital trials - use technology to implement clinical trials and collect data from people in their own home. The idea can be applied to the safety space. For example, wearables are useful for collecting data in locations, like on ships, that are difficult to access for researchers.
- A/B trials – where the technology itself is being investigated, where the conditions might be variations of an app or webpage. For example, testing the presentation of a digital risk assessment to see if a new version improves accessibility and potential downstream safety outcomes.

Beyond individual level trials

Thus far, trials have been mostly described in relation to the individual level. Individual level trials are particularly useful when you are interested in differences between participants or units and establishing effectiveness in a highly control design. However, tightly controlled individual designs can lack external validity. Group (cluster) level trial designs have better real-world utility in so far as design incorporates pre-existing groups which act as a unit, and where changes within the unit, for example a company, could lead to spill-over effects.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Cluster-based trials

Custer-based trials involve randomly allocating pre-existing groups, rather than individuals, to a given intervention. In the safety space this might be allocating construction companies (within a set of similar construction companies) to different conditions or allocating ships (within a fleet) to different conditions. Previously described designs (e.g. parallel, cross over and factorial) can be applied at the cluster level, however there are also specific types of cluster-level design.

A **stepped-wedge trial** is a specific type of cluster-based trial, where all clusters start off in the control condition and the intervention is incrementally rolled out to all clusters over time. A step-wedge trial includes randomisation by time and has the ethical advantage of enabling all clusters to access the intervention.

Split-plot trials

Split-plot trials include randomisation at more than one level and can include cluster and individual components.

- Whole plot randomisation level: involves cluster-based randomisation on elements that can be difficult to randomise. In the maritime example, where a ship-wide intervention would be advantageous in avoiding spillover effect, different ships would be randomised to different conditions.
- Subplot randomisation level: involves smaller units, clusters or individual level randomisation. In the maritime example, the crew on any given ship are randomly allocated to an intervention or not.

Split plot-trials can be complex and require careful planning but effective at assessing broad contextual effects as well as controlled individual ones in the same study.

Adaptive trials

Adaptive trials are not a design type per se but are increasingly popular in the medical field because they offer design flexibility based on the interim data. Adaptive methodology can be applied, and might be particularly useful, in the safety space where evidence for the effectiveness of one activity and not another, for example, found in the first part of the trial, is needed for the second part of the trial. It is paramount, however, that the adaptive elements are clearly laid out in the protocol at project initiation and that decisions made at the interim period are within the confines of the protocol and data driven.

Ethical considerations

With causal studies it is particularly important to anticipate, risk assess and mitigate for the principles of no harm and maintained data integrity. See chapter 3 for more information on:

- Voluntary informed consent
- Maximum benefit and no harm
- Fair opportunity to benefit
- Confidentiality, privacy and data protection
- Scientific integrity
- Social and environmental responsibility

Furthermore, where fair opportunity to benefit can be incorporated into any parallel group design studies, it should be. One way to do this is with a time randomised trial, which are encouraged if possible.

In medical trials concepts of sufficiently trained individuals, compliance and continued monitoring, are essential and interdependent pillars of ethical conduct. While trials in the safety space are less regulated (see chapter 1), these concepts should still be considered.

Sufficiently trained individuals

For all medical trials, staff curriculum vitaes are stored with the research documentation, which is internally and externally audited to ensure accountability for safety. Successful trials, irrelevant of field, need both experienced researchers and operational staff on board and it is important to consider partners for any trial projects. Should these not already be identified, the Global Safety Evidence Centre's team may be able to help, or provide expertise and advice. As part of the funding process at Lloyd's Register Foundation, organisational and professional credentials are requested. It is advantageous to show how methodological and operational trial elements will be overseen at this point.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Trial monitoring and compliance

While legal compliance is different in the safety space and medical fields given the different regulatory systems, data collected in a trial design should be continually monitored. In the safety space, monitoring is essential for the following reasons:

- To ensure compliance within the safety environment, i.e., that local and international standards and organisational policies are considered.
- To ensure continued compliance to the pre-registered study protocol.
- To ensure informed consent is considered on an ongoing basis but especially with any study changes or unexpected data.
- When investigating causality, it is possible that an intervention, predicted to improve outcomes, actually has negative effects which need to be identified at the earliest possible stage.

Early termination

It is often difficult to take away an intervention once it has been implemented. Staggered trial design, such as a step wedge design, can help with monitoring given the intervention is not rolled out all at once. Stopping criteria, i.e., thresholds or conditions, which when met mean discontinuation of data collection, should also be considered and in the protocol, to support the principle of no harm.

'Good' in relation to effective

Data integrity is a key part of trial ethics and should be planned for as rigorous designs do not avoid bias alone and powered studies are essential. Furthermore, significant causal differences do not necessarily mean 'good', but without rigour it is difficult to assess what 'good' is.

This is the catch-22 of experimental design in a developing field, where causal studies are required in order to show which measures are effective, but good measures are required for showing effectiveness. The best starting place is choosing those most established metrics and measures relevant to the context (see chapter 4). External validity is also important for establishing 'good'. Rigorously tested interventions might not be applicable in practice.

Research helps establish standards where they are currently unclear. Credible and robust must be considered alongside relevant and useful. However, a trial will not always be feasible! The idea is to generate the best possible evidence given the evidence base and resources. If you cannot run a trial, plan a case study or group comparison without randomisation.

Thinking about a trial? Some initial considerations

- Are there 'what works' research questions – experimental research is hypothesis driven – are improved safety outcomes expected based on the planned intervention(s) and existing evidence?
- Is the intervention(s) useful, relevant and suitable to what can be accommodated by the workers and organisation(s). It could be a single activity or programme of activities.
- Can existing safety activities be tested or will a new activity be trialled?
- What can be randomised, over time, units or groups?
- How is no harm to workers (or any units of study) ensured?
- Is there organisational infrastructure to run the trial? Are both researchers and senior management involved, to ensure feasibility and quality data?
- Writing a research proposal, protocol and analysis plan - what is the most suitable design?
- Ethical review – who will the ethics committee(s) be?
- Trial protocol pre-registration – where will it be submitted?
- Is there a data management and/or monitoring plan, to ensure ethical adherence and data integrity?

See chapter 2 for study plans, and contents.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Effectiveness and efficiency

For understanding 'what works' in term of risk protection and improvement of harm, RCTs are a vital tool, though are often overlooked because likelihood of harm differs drastically across different contexts and locations, safety and health are seen differently in society and because RCTs are seen as specific or resource intensive. The focus of this chapter was to highlight some options and considerations for running trials in the safety space to demonstrate effectiveness of interventions.

Another core component of what makes an intervention or activity useful is efficiency, which is a measure of how resource dependent or intensive that activity is. RCT methodology is strongly linked to cost-benefit analysis, as well as appraisal and evaluation methods more broadly. The relative cost of different safety activities is, of course, highly relevant to improving safety outcomes, particularly in industry. Assigning monetary values to safety outcomes can also aid cross-sector comparisons. These methods may be expanded on in future iterations of this guide. For now, some useful resources are here:

>> A practical framework for value driven grant making

>> Valuing Global Safety: Building a value for money framework

>> The Green Book - UK government guidance on appraisal

>> The Magenta Book - UK government guidance on evaluation

Further reading

Cochrane, A. L. (1972). Effectiveness and efficiency: Random reflections on health services. Nuffield Provincial Hospitals Trust.

Sanders, M., Ellingwood, J., Vallis, D., Rodriguez, G., Hume, S., & Kozman, E. (2025). Designing and delivering randomised trials for social policy. King's College London. https://www.kcl.ac.uk/sfg/assets/designing-and-delivering-randomised-trials-for-social-policy-digital.pdf

University of York. (n.d.). Directory of randomisation (randomization) software and services. https://www-users.york.ac.uk/~mb55/guide/randsery.htm

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., . . . Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. BMJ, 366, Article l4898. https://doi.org/10.1136/bmj.l4898

Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A.-W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., . . . Higgins, J. P. T. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. BMJ, 355, Article i4919. https://doi.org/10.1136/bmj.i4919

EGAP. (2024). Methods guides. https://egap.org/methods-guides/

Breckon, J., & Puttick, R. (2021). Quality in Qual: A proposed framework to commission, judge and generate good qualitative evaluation in wellbeing impacts. What Works Centre for Wellbeing. https://whatworkswellbeing.org/wp-content/uploads/2021/08/Quality-in-Qual-paper.pdf

The Turing Way Community. (2024). The Turing Way: A handbook for reproducible, ethical and collaborative data science. The Alan Turing Institute. https://book.the-turing-way.org/

Global Initiative for Industrial Safety. (n.d.). Making safety tech work: New framework guides adoption of industrial safety technologies. https://www.lrfoundation.org.uk/news/making-safety-tech-work-new-framework-guides-adoption-of-industrial-safety-technologies

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Chapter 8. Data management, storage, visualisation and reporting

Author: Aaron Gardner & Emily Lambe

How data is analysed, shared and stored matters as much to the integrity of a piece of safety research as the choice of method or test. Funders, journals and the wider research community increasingly expect that research outputs are transparent, reproducible and accessible beyond the lifetime of a project. In the safety space, where findings often feed into policy and operational decisions, openness is not an optional extra. This section sets out general principles that apply regardless of the analytical approach and is intentionally tool- and discipline-agnostic.

Pre-statement of hypotheses and analysis plans

As noted earlier in this guide, study intent (exploratory or hypothesis-testing) should be set apriori. The natural next step is to document the planned analysis before data is collected or accessed. A pre-specified analysis plan typically covers the primary and secondary hypotheses, the variables and how they are operationalised, the planned tests and their assumptions, the handling of missing data and outliers, and any planned subgroup or sensitivity analyses.

The point is not to constrain genuinely exploratory work, which has its own value, but to make the distinction between planned and post-hoc analyses visible to readers. Without that distinction, there is a real risk of running many tests and reporting only those that reach significance. This inflates false positive rates and links directly to the multiple comparisons issue covered earlier. Where deviations from the plan are needed (and they often are), they should be reported alongside the original plan rather than quietly substituted for it.

Pre-registration on a public registry, such as the Open Science Framework (OSF) or AsPredicted, formalises the process and provides a timestamped record. Pre-registration is increasingly available, and appropriate, for observational and secondary data work as well.

Open and reproducible analysis tooling

Where possible, analyses should be run using tools that allow the full chain from raw data to reported result to be inspected and re-run by someone else, and be accompanied by a statistical analysis plan. Open-source environments such as R and Python support this well, since analyses are captured as code rather than as a sequence of clicks. GUI-based tools such as Jamovi and JASP also support reproducible workflows and can be a sensible choice depending on the team. Proprietary tools are not ruled out but place a greater burden on the researcher to document each step in enough detail for others to follow.

Whichever tooling is chosen, analysis code should be version-controlled (for example using Git, hosted on GitHub or GitLab) and shared alongside the publication where data sensitivities allow. This makes errors easier to find and correct, supports independent validation, and lets methods be reused in later work.

Open data sharing

The default should be to share data openly, with appropriate exceptions for personal, commercially sensitive or safety-critical information. The FAIR principles (Findable, Accessible, Interoperable, Reusable) are a useful framework for what good data sharing looks like in practice and are widely referenced by funders.

Practical considerations include:

- Documenting the dataset with a clear data dictionary, codebook and provenance information, so future users understand what was collected, how, and what each variable means.
- Anonymising or aggregating data where individuals or organisations could otherwise be identified, taking into account jigsaw-identification risks where combining variables could re-identify a participant.
- Choosing a licence appropriate to the intended use (for example Creative Commons licences for documentation, more permissive licences for code).
- Sharing the analysis code, intermediate processed datasets, and any survey instruments or protocols, in addition to the raw data.

Where full open release is not possible, a tiered approach (public metadata with controlled access to the underlying data) is generally better than no sharing at all. Data sharing should be in accordance with informed consent and awareness of local and international laws and standards, such as General Data Protection Regulation (GDPR).

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Long-term storage

Local drives, personal cloud accounts and institutional servers tied to a single employer are not durable homes for research data. For long-term preservation, data and code should be deposited in a recognised repository that issues a persistent identifier (typically a DOI) and commits to retention over a defined period. Examples include Zenodo, Figshare, OSF and Dryad for general-purpose deposit, alongside domain-specific repositories where they exist.

Persistent identifiers matter because they allow the data and code to be cited directly, decouple access from the original hosting arrangement, and make it possible to link a published finding back to the exact version of the data and analysis that produced it.

Safety research often informs decisions that affect workers, the public and the design of systems. The credibility of those decisions rests on the credibility of the underlying evidence, and credibility is supported, not undermined, by letting the work be scrutinised. Pre-stated plans, open tooling, open data and durable storage together let findings be validated, challenged and built upon. That is how the evidence base improves.

Visualisation and communication of findings

The way findings are presented shapes how they are understood, remembered and acted on. In the safety space, where evidence often informs decisions made by non-researchers (policymakers, regulators, employers, workers and the public), poor visualisation or single-register reporting can limit the reach and usefulness of otherwise sound work. This section covers general principles for visualising and communicating quantitative findings and applies regardless of the tools used.

Visualisation principles

A good chart makes the underlying pattern in the data easier to see than the equivalent table or paragraph of text. A bad chart obscures it, or worse, misleads. Some general principles:

- The chart type should match the data type. Continuous data lends itself to scatter and line charts, categorical data to bars, and distributions to histograms, density plots or boxplots. Pie charts are rarely the best choice and should be avoided for more than a small number of categories.
- Axes should start at zero for bar charts and be clearly labelled, including units. Truncated axes can exaggerate small differences and are a common source of misleading visualisation.
- Colour should carry meaning, not decoration. Where colour encodes a variable, the choice of palette matters: sequential palettes for ordered data, diverging palettes for data with a meaningful midpoint, and qualitative palettes for unordered categories. Tools such as ColorBrewer and the viridis palette family are designed for this purpose.
- Uncertainty should be shown, not hidden. Error bars, confidence intervals and shaded ranges give readers a sense of how much weight to put on a result. A point estimate on its own can imply more precision than the data supports.
- Chart junk (heavy gridlines, 3D effects, drop shadows, unnecessary decoration) should be removed. The reader's attention should be on the data.

Labelling and annotation

Charts should be readable on their own, without the surrounding text. That means clear titles, axis labels, units, a legend where needed, and a caption or source line. Where a single feature in the chart carries the main message (a peak, a turning point, a notable outlier), a direct annotation on the chart is often more effective than relying on the reader to find it.

Direct labelling of series, where space allows, is usually clearer than a separate legend, since it removes the back-and-forth between chart and key. Sample sizes should be reported on or near the chart, particularly where groups differ in size.

Interactive visualisation

Interactive charts and dashboards can be useful where the audience benefits from exploring the data themselves: filtering by sector, drilling down by region, switching between time periods. Tools such as Tableau, Power BI, Observable, Plotly, Shiny (for R) and Streamlit (for Python) support this well.

Interactivity should add value rather than complexity. A static chart that answers the question clearly is preferable to an interactive one that requires the reader to assemble the answer themselves. Where interactivity is used, defaults should land on the most informative view, and the underlying data should be downloadable so users can take it away and work with it.

Interactive outputs also need a static fallback. PDFs, printed reports, screen-reader users and link-shared screenshots will not render the interactive version, so the key findings should also be available as static charts or summary tables.

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.

Accessibility

Accessibility is not an afterthought. Findings that cannot be read by part of the audience are findings that part of the audience cannot use. The Web Content Accessibility Guidelines (WCAG) set out the relevant standards, and the principles apply to research outputs as much as to public-facing websites.

Practical considerations include:

- Colour contrast between text and background, and between adjacent chart elements, should meet WCAG AA as a minimum. Online checkers make this easy to verify.
- Colour should not be the only channel carrying information. Patterns, shapes, direct labels or text annotations should also distinguish series, so that the chart remains readable in greyscale and for readers with colour vision deficiency. Around 8% of men and 0.5% of women have some form of colour vision deficiency, so this is not an edge case.
- Charts in digital outputs should have meaningful alternative text describing both what the chart shows and what the main finding is. "Bar chart of injury rates" is not useful alt text; "Bar chart showing injury rates by sector, with construction roughly three times the all-sector average" is.
- Tables should use proper table markup rather than being images of tables, so they can be read by screen readers and parsed by other tools.
- Plain-language summaries should sit alongside technical reporting, not replace it. Readability tools such as the Flesch-Kincaid score or the Hemingway Editor can help calibrate text to a target reading level.

Multiple levels of description for different audiences

A single research project will usually have several audiences with different needs: other researchers, funders, policymakers, practitioners, communicators, and the people whose work or safety the research concerns. One write-up cannot serve all of them well. A layered approach works better including:

- A headline finding of one or two sentences, suitable for a press release, a social media post or the opening of a briefing. This should be honest about uncertainty and avoid overclaiming.
- A plain-language briefing of one or two pages, written for a general audience, covering what was done, what was found, what it means in practice and what the main caveats are. Jargon should be defined or replaced.
- A technical summary for a policy or practitioner audience, going into method, sample, key results with effect sizes and confidence intervals, and limitations. Charts should be self-contained.
- A full technical report or paper for a research audience, including the analysis plan, full methods, all results (not only the significant ones), code and data references, and detailed limitations.

Each layer should be consistent with the others. A common failure mode is for the headline finding to drift away from what the technical report actually supports, often by dropping caveats or rounding uncertainty out of the picture. The plain-language and headline versions should be written, or at least signed off, by the research team rather than handed entirely to a communications function.

Safety findings often need to travel: from the analyst to a lead, from a regulator to an inspector, from a report to a toolbox talk. Each step is a point where the message can be lost, distorted or made unusable for the next audience. Clear visualisation, accessible formats and layered description reduce that loss and make it more likely that the evidence reaches the people whose decisions it is meant to inform.

Further reading

GOV.UK. (n.d.). Data protection. https://www.gov.uk/data-protection

GDPR-Info.eu. (n.d.). General Data Protection Regulation (GDPR). https://gdpr-info.eu/

Copyright © 2026 Lloyd's Register Foundation. All rights reserved.