

# Advancing Statistical Literacy in Eye Care: A Series for Enhanced Clinical Decision-Making

## Part 3: Design and Analysis of Clinical Research in Eye and Vision Studies

Daniela Oehring<sup>1,2</sup>

<sup>1</sup> Ph.D. · <sup>2</sup> Faculty of Health, University of Plymouth, Plymouth, UK

Received 29 December 2025; accepted 1 February 2026

### Abstract

**Purpose.** To provide comprehensive guidance on designing and analysing clinical research studies in ophthalmology and optometry, encompassing observational studies, non-randomised controlled trials, randomised controlled trials, and advanced trial methodologies including platform trials, adaptive designs, and decentralised trials. This article addresses the unique methodological challenges in eye and vision research while integrating contemporary frameworks such as the estimands approach for defining treatment effects and the ethical imperative of systematic reviews before initiating new trials.

**Material and Methods.** A comprehensive literature review was conducted using PubMed, Scopus, Web of Science, and Cochrane databases, focusing on study designs applicable to eye care research with emphasis on methodological innovations from 2019 to 2025. Simulated datasets representing common ophthalmic research scenarios were generated using Python to illustrate design principles and analytical approaches. International guidelines including ICH E9(R1), CONSORT 2025, and regulatory guidance from FDA and EMA were synthesised to provide current best-practice recommendations.

**Results.** Different study designs serve distinct purposes in advancing eye care knowledge. Observational studies excel at identifying associations and generating hypotheses but require careful consideration of confounding factors. Non-randomised controlled trials, strengthened by propensity score methods and target trial emulation frameworks, offer pragmatic evaluation when randomisation is not feasible.

Randomised controlled trials remain the gold standard for establishing causality but benefit from modern innovations including adaptive designs, platform trials enabling evaluation of multiple treatments, and the estimands framework clarifying precisely what treatment effects trials measure. Sequential Multiple Assignment Randomised Trials address the clinical reality that optimal treatment often requires adaptive sequencing. Decentralised trial elements expand access while maintaining scientific rigour. Key considerations for eye research include accounting for inter-eye correlation, managing paired organ data, and addressing masking challenges in surgical interventions.

**Conclusion.** Successful clinical research in eye and vision research requires matching study design to research questions, available resources, and ethical considerations. Modern trial methodology offers sophisticated tools for generating robust evidence efficiently, from the estimands framework ensuring trials measure clinically meaningful effects to platform designs enabling adaptive evaluation of multiple interventions. By understanding the strengths and limitations of each design type and applying appropriate analytical methods, researchers can advance evidence-based eye care while reducing research waste.

### Keywords

Clinical Research Design, Randomised Controlled Trials, Estimands Framework, Platform Trials, Adaptive Designs, Eye and Vision Research, Statistical Analysis, Inter-eye Correlation, Decentralised Clinical Trials, SMART Designs

## Introduction

Clinical research in ophthalmology and optometry has undergone a methodological revolution over the past decade, moving far beyond the traditional two-arm parallel randomised controlled trial. This transformation encompasses sophisticated study designs addressing complex questions about disease mechanisms, diagnostic accuracy, and treatment efficacy, while the unique anatomical and physiological characteristics of the visual system present both opportunities and challenges for researchers. This third instalment in our statistical literacy series builds upon the foundational concepts of descriptive statistics (Part 1)<sup>1</sup> and inferential testing (Part 2)<sup>2</sup> to explore the practical aspects of designing and analysing clinical studies in eye care, with particular emphasis on contemporary methodological innovations that can enhance research efficiency and validity.

The landscape of ophthalmic research encompasses diverse study designs, each suited to specific research questions and contexts. Observational studies have illuminated the natural history of diseases like age-related macular degeneration and identified risk factors for conditions such as myopic progression.<sup>3,4,5</sup> Non-randomised controlled trials have evaluated interventions in real-world settings where randomisation poses ethical or practical challenges. Randomised controlled trials have established the efficacy of treatments ranging from anti-VEGF therapy for retinal diseases to surgical techniques for glaucoma. However, the field now benefits from methodological advances that extend well beyond these traditional approaches, including platform trials that evaluate multiple treatments within shared infrastructure, adaptive designs that modify trial parameters based on accumulating evidence, and the estimands framework that ensures trials measure clinically meaningful treatment effects.

Despite advances in ophthalmic research, persistent methodological challenges remain. Reviews of eye and vision randomised controlled trials have documented frequent shortcomings in trial design and statistical analysis. These shortcomings include incomplete or inappropriate sample size and power calculations, and insufficient adjustment for correlated eye data in two-eye study designs. Such deficiencies undermine the validity of inferences.<sup>6</sup> These deficiencies not only waste resources but may lead to incorrect conclusions that affect patient care. The problem extends beyond individual studies: approximately 85% of research investment globally is estimated to be wasted due to asking irrelevant questions, flawed design, non-publication, and poor reporting.<sup>7</sup> In eye and vision research specifically, only 22.4% of phase III trials cite systematic reviews as justification for the study, missing opportunities to build on existing evidence.<sup>8</sup> Furthermore, the paired nature of eyes, the subjective nature of many visual outcomes, and the difficulty of masking surgical interventions create unique methodological considerations that distinguish eye research from other medical fields.

This article aims to equip clinicians and researchers with practical tools for designing robust studies and selecting appropriate analytical methods, integrating traditional approaches with contemporary innovations. We will explore

the continuum from observational to experimental designs, emphasising how research questions drive design choices. Special attention is given to challenges specific to eye research, including the statistical handling of bilateral organ data, the selection of clinically meaningful endpoints, and strategies for minimising bias in studies where traditional masking is impossible. Critically, we also address the ethical and scientific imperative of situating new trials within the context of existing evidence, the transformative potential of platform trials and adaptive designs, and the emerging role of decentralised trial elements in expanding access to clinical research.

## Material and Methods

The methods used in this article have been described in Part 1 of this series.<sup>1</sup> In brief, this article synthesises evidence from multiple sources to provide comprehensive guidance on clinical research design in eye and vision research. A narrative literature review was conducted using PubMed, Scopus, Web of Science, and the Cochrane Database of Systematic Reviews, covering publications from January 2019 to December 2025, with particular emphasis on methodological innovations published from 2019 onwards, as this period saw the finalisation of ICH E9(R1) on estimands (2019), updated CONSORT guidance (2025), and substantial advances in platform trial and adaptive design methodology. Search terms included combinations of clinical trial design, observational study, ophthalmology, optometry, statistical methods, inter-eye correlation, sample size, bias control, adaptive trials, platform trials, estimands, and decentralised trials.

Inclusion criteria encompassed methodological papers addressing study design in eye research, guidelines from regulatory bodies and professional organisations, systematic reviews of ophthalmic trials, landmark studies demonstrating exemplary design principles, and consensus statements on trial methodology. International regulatory guidance was reviewed including ICH E9(R1) on estimands, FDA guidance on adaptive designs and Bayesian methods, and EMA recommendations on decentralised trials. We excluded non-English publications, conference abstracts without full text, and studies focused solely on basic science research without clinical applications.

### The ethical and scientific foundation: Systematic Reviews before new trials

Before designing any new clinical study, researchers have both an ethical and scientific obligation to review existing evidence systematically. This principle, though seemingly obvious, remains inadequately implemented across medical research, contributing substantially to the estimated 85% of research investment that is wasted globally.<sup>7</sup> The 2014 Lancet series on increasing value and reducing waste in research established that research funders and regulators should demand that proposals be justified by systematic reviews of

existing evidence.<sup>8</sup> This requirement serves multiple purposes: it ensures that new research addresses genuine knowledge gaps, enables sample size calculations to be based on realistic effect estimates, and prevents unnecessary exposure of participants to inferior treatments when evidence already exists.

The consequences of ignoring existing evidence can be profound. A landmark cumulative meta-analysis by Lau et al. demonstrated that the mortality benefit of intravenous streptokinase for myocardial infarction could have been established by 1973, yet over a decade of additional trials enrolled patients to receive a placebo before the treatment became standard practice.<sup>9</sup> Similar delays likely occur in eye and vision research, though comprehensive cumulative meta-analyses are less common in our field. The principle that clinical trials should begin and end with systematic reviews reflects both the ethical imperative to minimise unnecessary research and the scientific value of situating new findings within the totality of evidence.<sup>10</sup>

Recent analysis demonstrates that despite growing awareness, implementation remains inadequate. Jia et al. examined trends in RCT reports citing prior systematic reviews from 2007 to 2021 and found that approximately 40% of RCT reports still fail to cite any systematic review, though this represents improvement from 65% a decade earlier.<sup>11</sup> In eye and vision research, the situation may be even more concerning, given the relatively young tradition of systematic evidence in the field. Practical implementation requires that researchers search for existing systematic reviews in their area before finalising protocols, conduct new systematic reviews when existing ones are outdated or unavailable, use quantitative synthesis to inform sample size calculations and choice of outcomes, and update systematic reviews to incorporate their new findings upon completion.

For eye and vision research specifically, researchers should consult the Cochrane Eyes and Vision Group reviews, search databases for existing meta-analyses of relevant interventions, and consider whether their proposed study addresses a genuine gap in evidence. When systematic reviews suggest equipoise truly exists, they provide the ethical justification for randomisation. When they reveal that evidence already favours one intervention, researchers must carefully consider whether additional trials are justified and, if so, what specific questions remain unanswered. In some circumstances, proceeding without a formal systematic review may be justified: when investigating entirely novel interventions or technologies with no comparable evidence (e.g., first-in-human trials of gene therapies for inherited retinal diseases), when addressing research questions in previously unstudied populations or settings, or when time-sensitive public health emergencies require rapid evidence generation. In such cases, researchers should document their rationale and conduct the most comprehensive literature search feasible within constraints.

## Fundamentals of study design in eye research

### The research question: foundation of design selection

Every successful clinical study begins with a clearly formulated research question. The quality and precision of this question determine not only the study design but also the

**Table 1:** Research question frameworks and their applications in ophthalmic research. Each framework provides a structured approach to formulating answerable clinical questions; PICO/PECO is most suitable for intervention and exposure studies, while SPIDER is designed for qualitative research. Researchers should select the framework that best matches their study type and research question.

Framework	Elements	Best Suited For	Example	Typical Design
PICO	Population, Intervention, Comparison, Outcome	Intervention efficacy/ effectiveness	Does intravitreal aflibercept improve visual acuity compared to laser for DMO?	RCT
PICOT	Population, Intervention, Comparison, Outcome, Time	Time-sensitive intervention questions	Does atropine 0.01% reduce myopia progression over 2 years compared to placebo?	RCT
PEO	Population, Exposure, Outcome	Aetiological/ risk factor questions	Does screen time exposure increase myopia incidence in children?	Cohort study
CoCoPop	Condition, Context, Population	Prevalence/ incidence estimation	What is the prevalence of glaucoma in adults over 40 in sub-Saharan Africa?	Cross-sectional
FINER	Feasible, Interesting, Novel, Ethical, Relevant	Evaluating any research question	Should we conduct this trial? Is it worth the investment?	All designs

eligibility criteria, sample size, endpoints, and analytical approach. Several frameworks assist researchers in developing answerable, clinically meaningful questions suitable for primary research, with the choice of framework depending on the type of question being asked, [Table 1](#).

### The PICO framework for intervention studies

The PICO framework (Population, Intervention, Comparison, Outcome), introduced by Richardson and colleagues in 1995, provides essential structure for intervention questions in clinical trials.<sup>12</sup> Each element maps directly to critical trial design decisions. The Population element defines who will be enrolled, translating into inclusion and exclusion criteria that balance internal validity against generalisability. The Intervention element specifies what treatment will be evaluated, including dose, duration, and mode of administration. The Comparison element identifies the control condition, whether placebo, active comparator, or usual care. The Outcome element determines what will be measured to assess treatment effects, guiding selection of primary and secondary endpoints.

In eye and vision research, the PICO framework requires careful consideration of the unique aspects of eye diseases and treatments. For Population, researchers must specify whether the study will include one or both eyes, the stage of disease, and relevant demographic factors. For Intervention, the route of administration (topical, intravitreal, systemic) and dosing regimen require explicit specification. For Comparison, the choice between placebo, sham procedure, or active comparator has both scientific and ethical implications. For Outcome, the selection must balance clinical relevance, measurement reliability, and sensitivity to change.

The PICOT extension adds Time as a fifth element, which proves essential when the timing of outcome assessment affects clinical interpretation. In a trial of anti-VEGF therapy for diabetic macular oedema, specifying that the primary outcome is mean change in best-corrected visual acuity at 12 months provides crucial information that shapes both the trial duration and the analytical approach. Similarly, a glaucoma trial might specify intraocular pressure reduction at 3 months as the primary endpoint while including visual field progression at 24 months as a secondary outcome.

### The PEO Framework for aetiological and prognostic studies

Not all clinical research questions involve interventions. When investigating risk factors, prognostic indicators, or disease associations, the PEO framework (Population, Exposure, Outcome) provides more appropriate structure than PICO. This framework, which emerged from epidemiological research traditions, replaces the Intervention and Comparison elements with Exposure, recognising that observational studies examine naturally occurring exposures rather than investigator-assigned treatments.

Consider the aetiological question: “Does high myopia increase the risk of retinal detachment?” Framing this using PEO yields Population (adults with refractive error), Exposure (high myopia, typically defined as spherical equivalent  $\leq -6.00$  D), and Outcome (retinal detachment). This structure naturally suggests a cohort study design comparing retinal detachment incidence between highly myopic and non-highly myopic individuals, with appropriate adjustment for confounding factors such as age, sex, and history of ocular trauma.

### The CoCoPop framework for prevalence and burden studies

Questions about disease frequency require yet another framework. The CoCoPop approach (Condition, Context, Population) structures prevalence and incidence questions where no exposure or intervention comparison exists. Developed within the Joanna Briggs Institute methodology, this framework helps researchers formulate precise descriptive questions.

For example, the question “What is the prevalence of diabetic retinopathy among Type 2 diabetics in urban primary care settings?” becomes Condition (diabetic retinopathy), Context (urban primary care settings), and Population (adults with Type 2 diabetes). This structure clarifies that a cross-sectional design with population-based sampling would be appropriate, and highlights the need for standardised diagnostic criteria and representative sampling within the specified context.

### The FINER criteria for research question evaluation

Before committing resources to a clinical trial, researchers should evaluate their question against the FINER criteria.<sup>13</sup> These criteria assess whether a research question merits the investment of time, funding, and participant involvement that clinical trials require.

**Feasibility** asks whether the study can realistically be completed. Can sufficient participants be recruited within a reasonable timeframe? Is the required expertise available? Are the measurements technically achievable? Is funding adequate? For ophthalmic trials, feasibility considerations include access to specialised equipment (OCT, visual field analysers, fundus cameras), availability of qualified outcome assessors, and the challenge of recruiting patients with specific disease stages.

**Interest** considers whether the question matters to the scientific community and will engage investigators sufficiently to complete a demanding study. A question that fails to generate enthusiasm among potential collaborators will struggle to achieve recruitment targets and maintain quality throughout follow-up.

**Novelty** asks whether the study will contribute new knowledge. This criterion connects directly to the ethical imperative

discussed earlier: systematic review of existing evidence should precede new trials to ensure genuine knowledge gaps exist. A trial that merely replicates well-established findings may waste resources and exposes participants to unnecessary risk.

**Ethics** requires that genuine equipoise exists regarding the treatments being compared and that the risks to participants are acceptable given potential benefits. In eye and vision research, where outcomes like vision loss can be irreversible, ethical considerations carry particular weight. Sham injection controls, while valuable for masking, require careful ethical justification.

**Relevance** asks whether the findings will influence clinical practice or health policy. A statistically significant difference that lacks clinical meaningfulness fails this criterion. In ophthalmic trials, clinical relevance often requires defining minimally important differences for outcomes like visual acuity (commonly 5–10 ETDRS letters) or intraocular pressure (typically 2–3 mmHg).

### Research question typology: Superiority, non-inferiority, and equivalence

The type of comparison intended fundamentally shapes trial design, sample size calculation, and statistical analysis. Researchers must explicitly specify whether they are asking a superiority, non-inferiority, or equivalence question, as each has distinct design implications.<sup>14</sup>

**Superiority questions** ask whether one treatment is better than another and represent the traditional hypothesis-testing framework. The null hypothesis states that treatments are equal, and the trial seeks evidence to reject this in favour of one treatment being superior. Sample size calculations for superiority trials depend on the minimum clinically important difference the investigators wish to detect, the expected variability, and the desired power.

**Non-inferiority questions** ask whether a new treatment is not unacceptably worse than an established standard. These questions arise when a new treatment offers advantages in cost, convenience, safety, or administration, even if efficacy is not superior. The null hypothesis states that the new treatment is inferior by more than a pre-specified margin, and the trial seeks to reject this hypothesis. Non-inferiority trials require careful specification of the non-inferiority margin, which should be smaller than the established treatment's effect versus placebo to ensure the new treatment remains superior to no treatment.

In eye and vision research, non-inferiority designs are increasingly relevant. A trial comparing a biosimilar anti-VEGF agent to the reference product might use a non-inferiority design, accepting equivalent efficacy while potentially offering cost advantages. Similarly, comparing a less invasive glaucoma procedure to trabeculectomy might employ non-inferiority testing, hypothesising that the new procedure provides similar IOP control with improved safety.

**Equivalence questions** ask whether two treatments produce essentially the same outcomes and require demonstrat-

ing that any difference falls within a pre-specified equivalence margin in both directions. These designs are less common in clinical practice but arise in bioequivalence studies and some comparative effectiveness contexts.

The distinction between these question types has profound implications for analysis and interpretation. Intention-to-treat analysis, which preserves the benefits of randomisation by analysing participants as randomised regardless of adherence, is conservative for superiority trials (biasing toward the null) but anti-conservative for non-inferiority trials (biasing toward equivalence). Consequently, non-inferiority trials should report both intention-to-treat and per-protocol analyses, with both required to support a non-inferiority conclusion.

## Hierarchy of evidence and study types

The evidence pyramid traditionally places systematic reviews and meta-analyses at the apex, followed by randomised controlled trials, cohort studies, case-control studies, and case series. However, this hierarchy assumes that the research question involves treatment efficacy and that traditional frequentist RCTs are the only option. Different questions require different optimal designs, and modern trial methodology offers alternatives that may be more efficient or ethical for specific contexts. Critically, the rigid hierarchy has been increasingly challenged as oversimplistic.<sup>15,16</sup> RCTs may be infeasible for rare diseases, unethical when effective treatments exist, or impractical for long-term outcomes such as cataract development over decades. In such circumstances, well-conducted observational studies with appropriate confounding control may provide the best available evidence. The key principle is matching the study design to the research question rather than reflexively privileging one design type. See [Table 2](#) and [Table 3](#).

### Observational studies versus experimentation

The most important division in clinical research methodology lies between observational and experimental studies. This distinction rests on a single question: does the investigator assign the exposure or intervention?

In observational research, the investigator examines exposures that occur naturally, such as smoking habits, dietary patterns, genetic variants, environmental conditions, or treatment choices made by clinicians and patients. Because these exposures are not randomly assigned, they may be associated with other factors (confounders) that independently affect the outcome, making causal inference challenging.

In experimental research, the investigator assigns the intervention, ideally using randomisation to ensure that treatment groups are comparable at baseline. This manipulation breaks the link between treatment assignment and potential confounders, enabling more confident causal conclusions.

**Table 2:** Study design selection matrix for clinical research in eye and vision research

Research Purpose	Optimal Design	Strengths	Limitations
Disease prevalence	Cross-sectional	Quick, cost-effective, population estimates	No temporal sequence, single time point
Risk factors	Cohort study	Temporal sequence clear, multiple outcomes	Time-consuming, expensive, attrition
Rare disease aetiology	Case-control	Efficient for rare outcomes, cost-effective	Recall bias, selection bias, temporal ambiguity
Treatment efficacy	Parallel RCT	Establishes causality, controls confounding	Costly, may lack generalisability
Multiple treatments	Platform trial	Shared infrastructure, adaptive features, efficient	Complex governance, statistical complexity
Treatment sequencing	SMART design	Builds dynamic treatment regimens, personalised	Complex analysis, larger sample size needed
Real-world effectiveness	Pragmatic trial	High external validity, practice-relevant	Lower internal validity, heterogeneity
Causal inference (RCT infeasible)	Target trial emulation	Uses existing data, addresses unmeasured bias	Requires rich data, residual confounding

However, experimental designs are not always feasible or ethical, necessitating observational approaches for many important questions.

**INFO BOX 1: Observational versus Experimental studies**

**Observational studies** are investigations in which the researcher observes and measures variables without manipulating them. Participants are not assigned to groups by the investigator; rather, they are observed in their natural state or as they have chosen (or been chosen by circumstance) to behave. The researcher’s role is to document what occurs, measure relevant variables, and analyse associations.

**Experimental studies** are investigations in which the researcher actively assigns participants to different conditions or interventions. The investigator controls who receives which treatment, enabling stronger conclusions about cause and effect. The randomised controlled trial is the most rigorous form of experimental study.

**The key question:** *Did the investigator decide which participants would receive which exposure or treatment?*

- If YES → Experimental study
- If NO → Observational study

**Temporal direction in research**

Studies vary in their temporal direction, timing of data collection, and analytical approach. Three primary designs – cross-sectional, cohort, and case-control studies – form the foundation of observational research. Landmark observational studies in eye and vision research have employed these designs to generate foundational knowledge: the Blue Mountains Eye Study, a population-based cohort study from Australia, has contributed substantially to our understanding of age-related macular degeneration risk factors and natural history<sup>3</sup>; the Collaborative Longitudinal Evaluation of Ethnicity and Refractive Error (CLEERE) Study has provided critical prospective data on myopia development in children<sup>4</sup>; and the Rotterdam Study has informed our understanding of genetic and environmental factors in eye disease.<sup>5</sup>

**INFO BOX 2: Temporal direction in research**

Study designs can be classified by their temporal orientation, the direction in which the investigation moves through time:

- **Cross-sectional studies** examine a single point in time. Exposure and outcome are measured simultaneously, providing a „snapshot“ of the population. Because both are measured at once, establishing which came first is impossible, limiting causal inference.
- **Longitudinal studies** follow participants over time, enabling observation of changes and temporal sequences. These include:

- i. *Prospective studies* begin in the present and follow participants forward into the future. The investigator defines the cohort, measures baseline characteristics, and waits for outcomes to occur.
- ii. *Retrospective studies* look backward from the present into the past. The investigator identifies individuals based on current status and examines historical records to determine prior exposures.

**Why temporal direction matters:** Establishing that an exposure preceded an outcome is essential for causal inference. Prospective designs provide the strongest evidence of temporal sequence because the exposure is documented before the outcome occurs.

### Cross-sectional studies

Cross-sectional studies provide a snapshot of disease prevalence and associated factors at a single time point. In eye and vision research, these studies have been instrumental in establishing the global burden of visual impairment and identifying populations at risk. Design considerations include sampling strategy, where population-based sampling provides the most generalisable results but requires substantial resources, while clinic-based sampling is more feasible but may introduce selection bias. Standardised measurements using validated instruments and standardised protocols ensure comparability across sites and studies; the Early Treatment Diabetic Retinopathy Study (ETDRS) chart has become the gold standard for visual acuity assessment. Quality control through training and certifying examiners, implementing duplicate grading for imaging, and monitoring inter-rater reliability is essential for data quality.

Statistical analysis of cross-sectional data requires prevalence estimates with confidence intervals accounting for the sampling design. When examining associations, logistic regression for binary outcomes or linear regression for con-

tinuous outcomes are commonly used, with adjustment for potential confounders. Complex survey designs require appropriate weighting and variance estimation methods.

### Cohort Studies

Cohort studies follow individuals over time to assess the development of outcomes, making them ideal for studying disease incidence and identifying risk factors. Prospective cohort studies are particularly valuable for understanding the natural history of eye diseases. Key design elements include cohort selection with clear inclusion and exclusion criteria, considering whether to use an inception cohort enrolled at disease onset or a prevalent cohort enrolled at various disease stages. Follow-up protocols should establish regular assessment intervals based on expected disease progression rates; more frequent assessments capture changes better but increase costs and participant burden. Retention strategies minimising attrition through participant engagement, flexible scheduling, and travel reimbursement are essential, and investigators should plan for 20-30% loss to follow-up in sample size calculations.

Statistical considerations for cohort data include survival analysis methods like Kaplan-Meier curves and Cox proportional hazards models that account for varying follow-up times. For repeated measurements, mixed-effects models or generalised estimating equations handle within-subject correlation appropriately.

### Case-Control studies

Case-control studies are particularly efficient for studying rare eye diseases or outcomes with long latency periods. By starting with the outcome and looking backwards for exposures, they require smaller sample sizes than cohort studies. Critical design decisions include case definition using strict

**Table 3:** Characteristics of major study designs in clinical research

Design	Selection	Direction	Strengths	Limitations	Eye and vision research use
<b>Cross-sectional</b>	Defined population at one time	Single time point	Quick, cost-effective, prevalence	No temporal sequence	AMD prevalence, refractive error surveys
<b>Prospective cohort</b>	Based on exposure status	Present to future	Clear temporal sequence, multiple outcomes	Time-consuming, expensive, attrition	Myopia progression, DR natural history
<b>Retrospective cohort</b>	Historical exposure groups	Past to present	Faster, uses existing data	Data quality limited by records	Long-term surgical outcomes
<b>Case-control</b>	Based on outcome status	Present to past	Efficient for rare outcomes	Recall bias, selection bias	Uveal melanoma risk factors
<b>RCT</b>	Random assignment	Present to future	Controls confounding, causality	Expensive, ethical constraints	DMO treatment, glaucoma surgery

diagnostic criteria to ensure cases truly have the condition of interest and considering whether to include incident or prevalent cases. Control selection requires that controls come from the same source population as cases and be at risk for developing the outcome; hospital controls may share healthcare-seeking behaviour with cases but might have other conditions affecting exposures. Matching on key confounders like age and sex can improve efficiency but requires matched analysis and prevents studying matched factors as risk factors.

**INFO BOX 3: How to identify study type from a publication**

When reading a clinical paper, ask these questions in sequence to identify the study design:

**Question 1: Did the investigators assign the treatment or exposure?**

- Yes → Experimental study (proceed to determine if randomised)
- No → Observational study (proceed to Question 2)

**Question 2: How were participants selected?**

- Selected based on exposure status → Cohort study
- Selected based on outcome status → Case-control study
- Selected from a defined population at one time point → Cross-sectional study

**Question 3: What is the temporal direction?**

- Follows participants forward from exposure to outcome → Prospective
- Looks backward from outcome to prior exposure → Retrospective
- Measures everything at one time point → Cross-sectional

**Common combinations:**

- Prospective cohort: Participants enrolled based on exposure, followed forward for outcomes
- Retrospective cohort: Historical records used to define past exposure groups and subsequent outcomes
- Case-control: Participants enrolled based on outcome, historical exposures compared
- Cross-sectional: Exposure and outcome measured simultaneously.

**Why randomisation matters: The problem of confounding**

The hierarchy of evidence places randomised controlled trials above observational studies primarily because randomisation addresses the fundamental problem of confounding. A confounder is a variable that is associated with both the exposure and the outcome, potentially creating a spurious association or masking a true one.

**INFO BOX 4: Understanding confounding**

**Confounding** occurs when a third variable distorts the apparent relationship between an exposure and an outcome.

**Example:** Suppose an observational study finds that patients who receive a new premium intraocular lens have better visual outcomes after cataract surgery than those receiving a standard lens. Before concluding the premium lens is superior, consider: patients receiving premium lenses may differ systematically from those receiving standard lenses. They may be younger, healthier, have less ocular comorbidity, or be treated by surgeons with greater expertise. Any of these factors could explain the better outcomes rather than the lens itself.

For a variable to be a confounder, it must:

- Be associated with the exposure (patients with certain characteristics are more likely to receive premium lenses)
- Be associated with the outcome (these same characteristics independently affect visual outcomes)
- Not be on the causal pathway between exposure and outcome (not be a consequence of receiving the premium lens)

**When observational designs are necessary or preferred**

Despite the advantages of randomisation, observational studies remain essential in clinical research for several reasons. First, many important questions cannot be addressed experimentally - we cannot randomise individuals to smoking or high myopia to study their effects on disease risk. Second, randomisation may be unethical when evidence already favours one treatment or when the intervention carries substantial risk. Third, observational studies reflect real-world practice, capturing effectiveness in routine clinical care rather than efficacy under controlled conditions. Fourth, observational designs are often more feasible for rare diseases, long-term outcomes, or hypothesis generation.

**Non-Randomised Controlled Trials and causal inference**

Non-randomised controlled trials, also known as quasi-experimental studies, evaluate interventions without random allocation. These designs are valuable when randomisation is impractical, unethical, or when evaluating health system interventions implemented at the population level. Modern methodological advances have substantially strengthened the causal inferences possible from non-randomised data.

## Before-After studies and interrupted time series

Before-after studies compare outcomes before and after an intervention within the same population. In eye and vision research, these are commonly used to evaluate quality improvement initiatives or the impact of new technologies. While feasible and pragmatic, these studies are vulnerable to temporal trends, regression to the mean, and the Hawthorne effect. Including a control group that does not receive the intervention strengthens causal inference.

### INFO BOX 5: Hawthorne effect

The Hawthorne effect refers to the phenomenon whereby individuals modify their behaviour in response to their awareness of being observed, rather than as a direct consequence of an experimental intervention. This effect can introduce systematic bias into research findings, particularly in clinical, behavioural, and social science studies, by inflating apparent intervention effects.<sup>17,18,19</sup> The Hawthorne effect originates from a series of industrial studies conducted at the Western Electric Hawthorne Works in Chicago during the late 1920s and early 1930s. The term itself was later coined and formalised based on these investigations.

Interrupted time series designs assess the impact of an intervention by analysing trends before and after implementation. Multiple pre- and post-intervention measurements help distinguish intervention effects from underlying trends. The analytical approach uses segmented regression analysis to model level change and slope change at the intervention point. Autoregressive integrated moving average (ARIMA) models can account for autocorrelation in the time series data, providing more valid standard errors and confidence intervals.

## Target trial emulation

When randomised trials are infeasible, target trial emulation offers a rigorous framework for causal inference from observational data. Developed by *Hernán and Robins*, this methodology specifies that observational analyses should explicitly emulate a hypothetical target trial by defining seven protocol components: eligibility criteria, treatment strategies, assignment procedures, follow-up period, outcome, causal contrast of interest, and analysis plan.<sup>20</sup> By designing observational analyses to mirror a well-designed RCT, researchers can identify and avoid common biases that afflict naive observational comparisons. In ophthalmology, target trial emulation could be applied to questions such as comparing anti-VEGF treatment regimens using electronic health record data, evaluating the long-term safety of refractive surgery using registry data, or assessing the effectiveness of different glaucoma treatment sequences in populations underrepresented in clinical trials.

The framework is particularly valuable for avoiding immortal time bias, a common error in observational studies where the period between cohort entry and treatment initiation is incorrectly attributed to the treatment group, artificially improving outcomes. *Hernán et al.* demonstrated how to apply the framework systematically, with subsequent regulatory acceptance of real-world evidence for certain decisions under the 21st Century Cures Act.<sup>21</sup> However, *Concato and Corrigan-Curay* emphasise that simply labelling data as ‘real-world’ does not specify study architecture or data quality; the rigour of the analytical approach determines validity.<sup>22</sup>

## Propensity Score methods

Propensity score methods attempt to balance confounders between treated and untreated groups in observational data, mimicking randomisation. The propensity score represents the probability of receiving treatment given observed covariates. Implementation strategies include matching, which pairs treated and untreated subjects with similar propensity scores; stratification, which divides subjects into strata based on propensity score quintiles; weighting, which uses inverse probability of treatment weighting in the analysis; and adjustment, which includes the propensity score as a covariate in regression models. Each approach has advantages and limitations, and the choice depends on the specific research context and data structure.

## Randomised Controlled Trials in eye care

Randomised controlled trials remain the gold standard for evaluating treatment efficacy, as randomisation balances both known and unknown confounders between groups. The rigour of RCTs has established many cornerstones of modern eye care, from anti-VEGF therapy for retinal diseases to surgical techniques for glaucoma. However, the traditional two-arm parallel RCT represents just one option within a diverse landscape of experimental designs, and modern methodology offers numerous refinements and alternatives.

## Randomisation methods

The randomisation method should ensure unpredictability while achieving balance between groups, **Table 4**. Simple randomisation gives each participant an equal probability of assignment to any group and works well for large samples but may result in imbalanced groups in smaller studies. Block randomisation ensures balance by randomising within blocks of predetermined size; for example, blocks of four with two treatments ensure equal numbers after every fourth participant. Stratified randomisation performs separate randomisation within strata of important prognostic factors, ensuring balance across these factors. Minimisation dynamically assigns

**Table 4: Randomisation methods: A practical guide**

Method	How it works	Advantages	Limitations	Best used when
<b>Simple randomisation</b>	Like flipping a coin: each participant has equal probability (e.g., 50%) of each group	Truly unpredictable; simple to implement; no restrictions on sequence	May produce unequal group sizes, especially in small trials (e.g., 70:30 split possible)	Large trials (n > 200) where imbalance probability is low
<b>Block randomisation</b>	Randomises within small blocks (e.g., block of 4: ABBA, ABAB). Groups balanced after each complete block	Guarantees equal or near-equal group sizes; balance maintained throughout recruitment	Last assignment in block may be predictable; requires variable block sizes to maintain concealment	Most trials; essential for small-to-medium studies (n < 200)
<b>Stratified randomisation</b>	Separate block randomisation within strata (e.g., by site, disease severity, age group)	Ensures balance on key prognostic factors; prevents chance imbalances in important subgroups	Limited to 2-3 stratification factors; many strata create small cells; requires advance planning	Multi-centre trials (stratify by site); known strong prognostic factors
<b>Minimisation</b>	Algorithm assigns each participant to the group that minimises overall imbalance across multiple factors	Balances many factors simultaneously; works well in small trials; adapts dynamically	Assignments may be predictable; requires computer system; some consider it 'quasi-random'	Small trials with multiple important prognostic factors (> 3 factors)
<b>Cluster randomisation</b>	Randomises groups (clinics, practices, communities) rather than individuals	Essential when intervention applies at group level; avoids contamination between groups	Requires more participants (design effect); analysis must account for clustering; fewer randomisation units	Screening programme evaluations; practice-level interventions; policy studies

participants to minimise imbalance across multiple factors simultaneously and is particularly useful for smaller trials with multiple important covariates.

### Allocation concealment and blinding

Allocation concealment prevents selection bias by ensuring that investigators cannot predict or influence group assignment. Methods include central randomisation systems (web-based or telephone), sequentially numbered opaque sealed envelopes, and pharmacy-controlled randomisation for drug trials. Blinding (masking) prevents performance and detection bias, but eye and vision research presents unique challenges. For surgical interventions, true sham surgery is often unethical, necessitating masked outcome assessors and standardised post-operative care protocols. For device studies, different device appearances may preclude patient blinding, so efforts should focus on blinding outcome assessors and data analysts. For injection studies, sham injections can maintain blinding but require ethical justification; masked visual acuity examiners and reading centres for imaging outcomes provide alternative approaches.

#### INFO BOX 6: Randomisation versus allocation concealment – Two different concepts

These terms are often confused but refer to distinct safeguards against bias:

**Randomisation** is the method used to generate the assignment sequence. It determines WHO gets WHICH treatment. Good randomisation ensures the sequence is genuinely unpredictable (e.g., computer-generated random numbers, not alternation or birth dates).

**Allocation concealment** is the method used to HIDE the upcoming assignment from those enrolling participants. It prevents investigators from knowing future assignments and selectively enrolling patients. Even perfect randomisation is undermined if investigators can see the next assignment.

**Why both matter:**

- A trial with good randomisation but poor concealment: Investigator sees the next patient will receive the experimental drug and delays enrolling a sicker patient until a placebo slot appears → Selection bias
- A trial with poor randomisation but good concealment: Alternating assignment (odd/even) hidden by sealed envelopes → Still potentially predictable if the pattern is discovered

**Gold standard for allocation concealment:**

- Central randomisation (web-based or telephone system)
- Pharmacy-controlled drug dispensing
- Sequentially numbered, opaque, sealed envelopes (acceptable but less secure)

**Remember:** Randomisation creates comparable groups; allocation concealment protects the randomisation from subversion.

## Sample size calculation

Adequate sample size ensures sufficient power to detect clinically meaningful differences while avoiding waste of resources. Key parameters include Type I error ( $\alpha$ ), usually set at 0.05 with consideration of adjustment for multiple endpoints or interim analyses; power ( $1-\beta$ ), typically 80-90% with higher power for pivotal trials; effect size, defined as the clinically meaningful difference (e.g., 5 ETDRS letters for visual acuity); variance estimated from pilot studies or literature; inter-eye correlation if using both eyes; and anticipated dropout rate to inflate sample size appropriately.

For studies using both eyes, the effective sample size is reduced due to correlation. The design effect can be calculated as:  $\text{Design Effect} = 1 + (m - 1) \times \rho$ , where  $m$  is the number of observations per subject (2 for both eyes), and  $\rho$  is the inter-eye correlation coefficient, typically ranging from 0.3 to 0.7 for many ophthalmic outcomes. For example, if a study initially calculates that 100 subjects (200 eyes) are needed assuming independence, and the true inter-eye correlation is  $\rho = 0.5$ , the design effect =  $1 + (2 - 1) \times 0.5 = 1.5$ . The effective sample size is therefore  $200 / 1.5 = 133$  independent units, requiring approximately 150 subjects (300 eyes) to achieve the originally planned power. **Table 5** summarises key parameters for sample size calculation in ophthalmic trials.

**Table 5:** Key parameters for sample size calculation in ophthalmic trials

Parameter	Considerations for Eye Research
<b>Type I error (<math>\alpha</math>)</b>	Usually 0.05; adjust for multiple endpoints, interim analyses, or adaptive designs
<b>Power (<math>1-\beta</math>)</b>	Typically 80-90%; higher power for pivotal regulatory trials
<b>Effect size</b>	Define clinically meaningful difference (e.g., 5-10 ETDRS letters, 2-3 mmHg IOP)
<b>Variance</b>	Estimate from pilot data or systematic reviews; account for measurement variability
<b>Inter-eye correlation</b>	If using both eyes, adjust for correlation (typically $\rho = 0.3-0.7$ )
<b>Dropout rate</b>	Inflate sample size by expected attrition (10-30% depending on follow-up)

## The Estimands Framework: Defining what trials measure

The ICH E9(R1) addendum, effective since 2020, introduced the estimands framework – a structured approach to defining precisely what treatment effect a trial intends to estimate.<sup>23</sup> This framework represents a fundamental shift in how clinical trials are designed, conducted, and analysed, ensuring that the research question, trial design, and statistical analysis are explicitly aligned. For eye and vision research, where treatment discontinuation, rescue therapy, and switching are common, the estimands framework has particular relevance.

### The five attributes of an Estimand

An estimand has five defining attributes that together specify exactly what treatment effect is being estimated.

1. The **treatment** attribute specifies the intervention being evaluated, including dose, duration, and any relevant aspects of administration.
2. The **population** attribute defines who the treatment effect applies to, including inclusion and exclusion criteria and any restrictions on generalisability.
3. The **variable** attribute identifies the endpoint or outcome of interest, including how and when it is measured.
4. The **intercurrent events** attribute addresses how post-randomisation events that affect interpretation will be handled.
5. Finally, the **population-level summary** attribute specifies the statistical measure of effect, such as difference in means, hazard ratio, or odds ratio.

The concept of intercurrent events is central to the framework. These are post-randomisation occurrences that affect either the interpretation or the existence of outcomes,

including treatment discontinuation, rescue medication use, treatment switching, and death. Rather than treating these as missing data problems to be handled in the analysis phase, the estimands framework requires explicit decisions about how each type of intercurrent event relates to the scientific question of interest.

### Strategies for handling intercurrent events

Five strategies exist for handling intercurrent events, each answering a different clinical question.

1. The **treatment policy strategy** includes all outcomes regardless of intercurrent events, essentially answering ‘What happens if we assign this treatment?’ regardless of what patients actually do. This approach aligns with traditional intention-to-treat analysis.
2. The **hypothetical strategy** estimates the treatment effect that would have been observed if the intercurrent event had not occurred, answering questions like ‘What would the effect be if patients had not discontinued treatment?’
3. The **composite strategy** incorporates the intercurrent event into the endpoint definition, such as defining the outcome as ‘vision improvement without requiring rescue therapy.’
4. The **while-on-treatment strategy** includes only outcomes while on the randomised treatment, useful when the effect of actual treatment receipt is of interest.
5. The **principal stratum strategy** focuses on subsets of patients defined by their potential intercurrent event behaviour, such as ‘patients who would adhere to treatment regardless of assignment.’

### Application to eye and vision research

The practical implications for ophthalmic trials are substantial. *Kahan et al.* provided an accessible primer using worked examples that clinicians can adapt<sup>24</sup>, while *Ratitch et al.* offered detailed guidance on defining efficacy estimands.<sup>25</sup> Consider a trial of a new anti-VEGF agent for diabetic macular oedema where patients may discontinue the study drug and receive rescue treatment with an alternative anti-VEGF agent. Different estimands answer different questions:

- A treatment policy estimand asks about the effect of initiating the new agent versus comparator regardless of subsequent treatment changes, which may be most relevant for health policy decisions.
- A hypothetical estimand asks what the effect would be if no one received rescue therapy, though this may be unrealistic clinically.
- A composite estimand might define success as visual improvement without requiring rescue therapy, capturing both efficacy and tolerability.

In glaucoma trials, where medication discontinuation and switching are common, the estimands framework helps clarify whether the trial aims to assess the effect of the initial treatment assignment (treatment policy), the effect if patients had maintained their assigned treatment (hypothetical), or the effect while actually receiving the assigned medication (while-on-treatment). Each answers a legitimate but distinct clinical question, and the choice should be driven by the intended use of the evidence. Fletcher and colleagues documented implementation experience marking two years of applying the new framework, providing practical guidance for researchers transitioning to estimand-based trial design.<sup>26</sup>

### Advanced trial designs

The traditional two-arm parallel RCT, while rigorous, is not always the most efficient or informative design for addressing clinical questions. Modern trial methodology offers sophisticated alternatives that can improve efficiency, enable adaptive learning, and better address the complexity of real-world treatment decisions.

#### Crossover and Factorial designs

**Crossover trials** have each participant serve as their own control by receiving all treatments in sequence. This design is particularly powerful for chronic conditions with reversible treatment effects, offering reduced sample size requirements (typically 25–50% of parallel design), control for between-subject variability, and particular suitability for comparing eye drops or contact lenses. However, limitations include the requirement for adequate washout periods to avoid carryover effects, unsuitability for curative treatments or progressive diseases, and potential period effects if disease state changes over time.

**Factorial designs** test multiple interventions simultaneously, allowing assessment of individual effects and interactions. A 2 × 2 factorial design tests two interventions, each at two levels. For example, testing both a new surgical technique and post-operative medication in cataract surgery creates four groups: standard surgery plus placebo, standard surgery plus medication, new surgery plus placebo, and new surgery plus medication. This design efficiently evaluates two questions within a single trial while also assessing whether the interventions interact.

#### Platform trials and master protocols

The COVID-19 pandemic demonstrated the extraordinary power of platform trials – perpetual infrastructure that can evaluate multiple interventions simultaneously against shared control arms. The RECOVERY trial, which enrolled over 45,000 patients across 176 UK NHS hospitals, established dexamethasone as the first treatment to reduce COVID-19 mortality, fundamentally changing practice within months

of initiation.<sup>27</sup> *Woodcock and LaVange's* seminal paper from the FDA established the taxonomy distinguishing platform, basket, and umbrella trials under the 'master protocol' framework.<sup>28</sup>

Platform trials offer several key advantages over traditional RCTs: shared infrastructure reduces per-comparison costs; adaptive features allow ineffective treatments to be dropped while adding new treatments as they become available; Bayesian methods enable continuous learning and more intuitive probability statements; and shared control arms improve efficiency while maintaining scientific rigour. The *Adaptive Platform Trials Coalition* provided comprehensive recommendations for design, conduct, and reporting<sup>29</sup>, while *Berry et al.* articulated the efficiency arguments favouring platform approaches.<sup>30</sup>

For eye and vision research, platform trials offer particular promise for conditions where multiple treatments are under investigation. A platform trial for diabetic macular oedema could simultaneously evaluate multiple anti-VEGF agents, different dosing regimens, and combination approaches against a shared standard-of-care arm. Similarly, platform infrastructure for glaucoma device trials could efficiently compare multiple minimally invasive glaucoma surgery devices while new technologies continue entering the market. *Park et al.'s* systematic review identified 83 master protocol trials through 2019, demonstrating growing adoption of these approaches.<sup>31</sup>

## Sequential Multiple Assignment Randomised Trials (SMARTs)

Sequential Multiple Assignment Randomised Trials address a fundamental clinical reality: optimal treatment often requires adaptive sequencing based on patient response. Unlike traditional RCTs that test fixed interventions, SMARTs randomise patients at multiple decision points to build evidence for dynamic treatment regimens. *Susan Murphy's* foundational 2005 paper introduced the theoretical framework, demonstrating that sequential randomisation is necessary because past treatment effects may have delayed consequences that influence subsequent treatment choices.<sup>32</sup>

The methodology has since been refined through accessible tutorials by *Almirall* and *Nahum-Shani*, who provided practical guidance on design, power analysis, and longitudinal data analysis.<sup>33,34</sup> SMARTs are particularly appropriate for chronic disease management where treatment must adapt over time. In eye and vision research, myopia progression management presents an ideal application: children might be initially randomised to atropine versus orthokeratology, then re-randomised based on response to continue current treatment versus switch or augment with the alternative approach. *Artman et al.* addressed sample size considerations specifically for comparing embedded dynamic treatment regimens<sup>35</sup>, while *Nahum-Shani et al.* provided tutorials for longitudinal analysis of SMART data.<sup>36</sup>

## Bayesian adaptive designs

Bayesian methods provide clinicians with the probability that a treatment works given observed data – a more intuitive interpretation than frequentist p-values. The framework incorporates prior information, updates beliefs as data accumulate, and enables predictive probability calculations for futility and success. The FDA's 2010 guidance on the use of Bayesian statistics in medical device clinical trials remains the foundational regulatory document for these methods, with subsequent international methodological reinforcement provided through ICH E9(R1), underscoring their relevance for ophthalmic device evaluation.<sup>23,37</sup>

The key insight is that 'today's posterior becomes tomorrow's prior' – Bayesian learning is inherently sequential. This enables efficient interim analyses without the rigid stopping rules of frequentist approaches. Rather than asking whether we can reject the null hypothesis at a predetermined alpha level, Bayesian analyses ask what probability we should assign to meaningful treatment effects given all accumulated evidence. *Muehleman and colleagues' 2023 tutorial* specifically addresses the knowledge gap that limits Bayesian adoption, explaining concepts starting from familiar clinical practice and using the Pfizer/BioNTech COVID-19 vaccine trial as a detailed case study.<sup>38</sup>

Adaptive trials allow pre-specified modifications based on accumulating data without compromising validity. Common adaptations include sample size re-estimation based on observed variance, dropping ineffective treatment arms, enrichment strategies focusing on responsive subgroups, and seamless phase II/III designs. For rare eye diseases where traditional trials would require impractically long enrolment periods, Bayesian adaptive approaches enable efficient evidence generation with smaller sample sizes while maintaining appropriate uncertainty quantification.

When choosing between frequentist and Bayesian approaches, clinicians should consider several practical factors. Frequentist methods offer familiarity, regulatory precedent, and straightforward interpretation of confidence intervals, making them suitable when prior information is limited or controversial. Bayesian methods excel when incorporating prior knowledge is appropriate (e.g., from earlier trial phases), when continuous monitoring with flexible stopping is desired, when sample sizes are small (as in rare eye diseases), or when probability statements about treatment effects are more clinically intuitive than hypothesis testing. The choice is not mutually exclusive: many modern trials use Bayesian methods for interim decisions while reporting frequentist statistics for regulatory submissions.

## Decentralised clinical trials

COVID-19 forced rapid adoption of decentralised clinical trial (DCT) elements, with over 500 DCTs conducted annually since 2020. These trials replace or supplement traditional site visits with remote data collection via telehealth, wearables, electronic patient-reported outcomes, and home nursing

visits. The FDA’s September 2024 final guidance and EMA’s 2022 recommendations establish regulatory frameworks emphasising patient safety and data integrity.<sup>39,40</sup> Park et al. provided a comprehensive landscape analysis of DCT development and regulatory guidance.<sup>41</sup>

Electronic diaries and electronic clinical outcome assessments show highest current adoption (51–56 % across major pharmaceutical companies), while local physician networks and telemedicine show the greatest growth potential.<sup>42</sup> Digital endpoints require rigorous ‘fit-for-purpose’ validation covering technological validity and clinical validation including repeatability, patient/control differentiation, and correlation with traditional endpoints. *Kruizinga* et al. developed a structured framework for validating novel digital endpoints.<sup>43</sup>

For eye and vision research, DCT elements present both opportunities and limitations. Home visual acuity testing using validated smartphone applications shows promise for monitoring trials, as do remote fundus photography systems and visual function questionnaires administered electronically. Patient-reported outcome measures for symptoms and quality of life are particularly amenable to remote collection. However, core ophthalmic assessments including slit-lamp examination, Goldmann applanation tonometry, optical coherence tomography, and automated perimetry remain fundamentally site-based due to equipment requirements and the need for trained operators.

Hybrid designs combining DCT elements with essential site visits appear optimal for most ophthalmic research. For example, a dry eye trial might conduct screening and periodic safety assessments at sites while collecting daily symptom diaries and quality-of-life measures remotely. This approach expands geographic access, reduces participant burden, and enables more frequent outcome assessment while maintaining the specialised examinations essential for eye care research. The Clinical Trials Transformation Initiative’s foundational recommendations, which predated the pandemic, have shaped subsequent regulatory guidance and provide a framework for implementing DCT elements responsibly.<sup>42</sup>

## Special considerations for eye and vision research

### Managing paired organ data

The paired nature of eyes presents unique statistical challenges that distinguish ophthalmic research from most other medical fields. Different approaches are appropriate depending on the research question and study design. Selecting one eye per subject eliminates correlation issues but reduces power; selection methods include random selection, choosing the worse or better eye based on disease severity, or using the first eye to meet eligibility criteria. Including both eyes increases power but requires appropriate statistical methods to account for the correlation: mixed-effects models include random effects for subjects, generalised estimating equations spec-

ify correlation structure between eyes, and robust standard errors adjust for clustering within subjects.

Paired eye designs assigning different treatments to each eye of the same patient provide perfect matching but require careful consideration. Systemic effects may contaminate the control eye, introducing bias. Patient preference or psychological effects may differ when both treatments are experienced simultaneously. These designs are most appropriate for localised treatments without systemic absorption, such as comparing two different contact lens designs.

### INFO BOX 7: The unit of analysis problem in eye research

Ophthalmic research faces a unique methodological challenge: most people have two eyes, but most statistical methods assume independent observations.

**Why this matters:** If a study includes both eyes from each participant, the observations are not independent – eyes from the same person tend to be more similar to each other than to eyes from different people. Analysing both eyes as if they were independent observations artificially inflates the sample size and underestimates standard errors, leading to falsely narrow confidence intervals and inflated Type I error rates.

#### Solutions include:

- One eye per participant: Eliminates the correlation problem but reduces statistical power
- Mixed-effects models: Include a random effect for participant, explicitly modelling the within-person correlation
- Generalised estimating equations (GEE): Account for clustering of eyes within participants
- Paired-eye designs: Use within-person comparison when different treatments can be assigned to each eye

**The key principle:** The statistical method must match the unit of analysis. If both eyes contribute data, the analysis must account for their correlation.

### Selecting clinically meaningful endpoints

Choosing appropriate primary outcomes is crucial for trial success and clinical relevance. Endpoints should be clinically relevant and meaningful to patients and practitioners, reliable and valid with reproducible measurement properties, sensitive to change and able to detect treatment effects, and feasible to measure in the study setting. [Table 6](#) summarises common primary endpoints across major ophthalmic disease areas.

**Table 6:** Common primary endpoints in ophthalmic clinical trials by disease area\*

Disease Area	Common primary endpoints	Considerations
<b>AMD</b>	Mean change in BCVA; proportion gaining ≥15 letters	Ceiling effects in good vision; consider low luminance VA
<b>Glaucoma</b>	IOP reduction; visual field progression	IOP as surrogate; VF requires long follow-up
<b>Diabetic Retinopathy</b>	DR progression; macular thickness	Standardised grading scales; OCT variability
<b>Dry Eye</b>	Symptom scores; tear film metrics	Poor correlation between signs and symptoms
<b>Myopia Control</b>	Axial length progression; refractive error change	Axial length more reliable than refraction

\* Minimum clinically important differences (MCIDs) vary by endpoint: for visual acuity, 5-10 ETDRS letters is typically considered clinically meaningful; for intraocular pressure, 15-20% reduction is often used as a threshold; for patient-reported outcomes, half a standard deviation is a common rule of thumb. Researchers should justify their chosen effect sizes based on published MCIDs and clinical relevance.

### Pragmatic trials and real-world evidence

Pragmatic trials prioritise external validity and real-world applicability, in contrast to explanatory trials focused on controlled efficacy demonstration. Ford and Norrie provided a comprehensive overview distinguishing these approaches<sup>44</sup>, emphasising that the choice between pragmatic and explanatory designs depends on the clinical question being addressed. The PRECIS-2 tool provides a nine-domain ‘wheel’ helping trialists explicitly design trials matching their intended purpose, with each domain scored from 1 (very explanatory) to 5 (very pragmatic).<sup>45</sup>

The nine PRECIS-2 domains assess eligibility criteria, recruitment path, setting, organisation, flexibility in delivery, flexibility in adherence, follow-up, primary outcome, and primary analysis. By explicitly scoring each domain, researchers can identify whether their design matches their intended use of the results. A highly explanatory trial with strict eligibility, intensive monitoring, and per-protocol analysis may demonstrate whether an intervention can work under ideal conditions, but results may not generalise to routine practice. A pragmatic trial with broad eligibility, usual care comparators, and intention-to-treat analysis addresses whether an intervention does work in practice.

For eye and vision research, pragmatic trials are particularly valuable for comparing surgical techniques where efficacy is established but comparative effectiveness in routine practice is unknown, evaluating screening programmes where implementation factors strongly influence outcomes, and assessing care pathways where the intervention is a system-level change rather than a discrete treatment. Registry-based randomised trials represent an emerging hybrid approach where randomisation is embedded within existing clinical registries, enabling large-scale pragmatic trials with minimal additional infrastructure.

### Statistical analysis of clinical research data

#### Analysis principles

The intention-to-treat (ITT) principle analyses participants as randomised regardless of adherence, preserving randomisation benefits and providing conservative efficacy estimates. Per-protocol analysis includes only participants who completed the protocol as planned, potentially overestimating treatment effects but useful for understanding efficacy under ideal conditions. Best practice involves conducting both analyses, with ITT as the primary analysis for superiority trials. For non-inferiority trials, both ITT and per-protocol analyses should demonstrate non-inferiority, as ITT alone can bias towards equivalence. The estimands framework clarifies how these choices relate to the scientific question of interest.

Missing data are inevitable in clinical trials and can bias results if not handled appropriately. Modern approaches include multiple imputation, which creates multiple plausible values for missing data based on observed patterns; mixed models, which use all available data under missing at random assumptions; and sensitivity analyses, which test robustness under different missing data assumptions including missing not at random scenarios.

#### Modern multiplicity methods

Clinical trials often involve multiple endpoints, time points, subgroups, and treatment comparisons, increasing the risk of Type I errors. Modern multiplicity methods have advanced far beyond simple Bonferroni correction. Bretz et al. introduced the graphical approach to multiple testing procedures, representing multiple test procedures as directed weighted graphs where nodes correspond to hypotheses.<sup>46,47</sup> This visualisation enables clinicians and regulators to understand complex testing hierarchies intuitively.

**INFO BOX 8: Understanding the Bretz Graphical Approach to Multiple Testing**

The graphical approach represents hypotheses as nodes in a graph, with directed edges showing how significance level ( $\alpha$ ) is redistributed when hypotheses are rejected.

**Key components:**

- Nodes: Each hypothesis ( $H_1, H_2, \text{etc.}$ ) is a circle showing its current  $\alpha$  allocation
- Edges: Arrows between nodes with weights (summing to 1 for outgoing edges) showing how  $\alpha$  flows when a hypothesis is rejected
- Initial allocation: Total  $\alpha$  (usually 0.05) divided among hypotheses based on clinical priorities

**The algorithm:**

1. Test each hypothesis at its current  $\alpha$  level
2. If a hypothesis is rejected ( $p < \alpha$ ), remove its node from the graph
3. Redistribute its  $\alpha$  to connected hypotheses according to edge weights
4. Update the graph and repeat until no more hypotheses can be rejected

**Why it matters:** Unlike simple Bonferroni (which divides  $\alpha$  equally and wastes ‘unused’  $\alpha$  when some hypotheses are rejected), the graphical approach recycles  $\alpha$  to remaining hypotheses, increasing power while still controlling the family-wise error rate.

Worked example: Anti-VEGF Dose-Finding Trial (see Figure 1). Consider a phase III trial comparing high-dose and low-dose anti-VEGF against placebo for wet AMD, with co-primary endpoints of visual acuity (VA) and central retinal

thickness (CRT). This creates four hypotheses:  $H_1$  (high dose vs placebo, VA),  $H_2$  (low dose vs placebo, VA),  $H_3$  (high dose vs placebo, CRT), and  $H_4$  (low dose vs placebo, CRT).

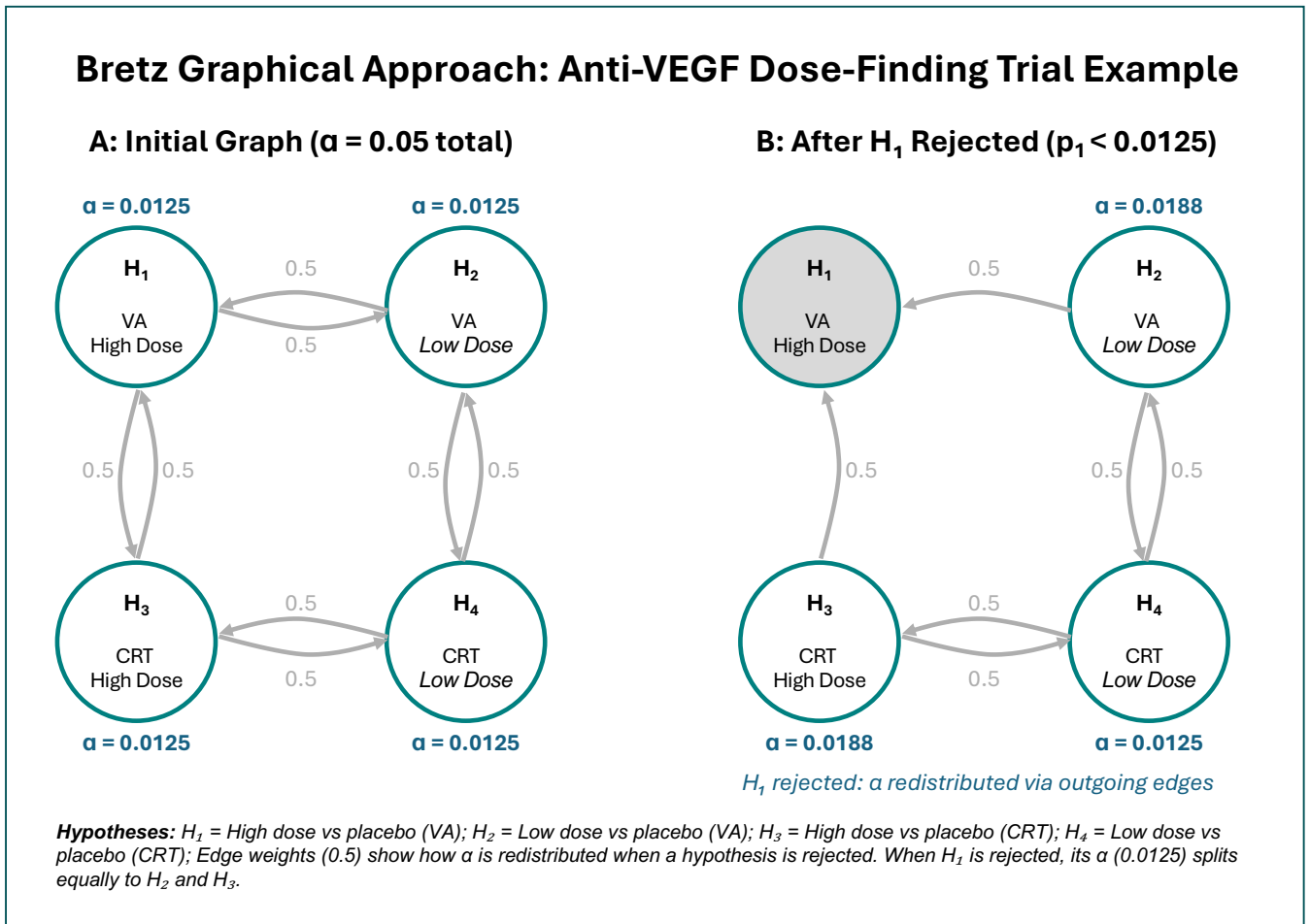
In the initial graph (Figure 1A), the total  $\alpha = 0.05$  is divided equally among all four hypotheses ( $\alpha = 0.0125$  each). The edge weights of 0.5 indicate that when any hypothesis is rejected, half of its  $\alpha$  goes to the hypothesis testing the same dose for the other endpoint, and half goes to the hypothesis testing the other dose for the same endpoint. This structure reflects the clinical logic that evidence for one dose informs our confidence in testing both endpoints for that dose, and that evidence for one endpoint informs testing both doses for that endpoint.

Suppose the trial finds  $p_1 = 0.008$  for the high-dose VA comparison ( $H_1$ ). Since  $0.008 < 0.0125$ , we reject  $H_1$ . The graph updates (Figure 1B):  $H_1$  is removed, and its  $\alpha = 0.0125$  is redistributed according to the outgoing edge weights.  $H_2$  (low dose VA) receives  $0.0125 \times 0.5 = 0.00625$ , bringing its total to 0.01875.  $H_3$  (high dose CRT) also receives 0.00625, bringing its total to 0.01875.  $H_4$  remains at 0.0125. The procedure continues with the updated  $\alpha$  levels, providing more power to test the remaining hypotheses than would be available under Bonferroni correction.

The graphical approach has been extended to incorporate weighted *Simes* tests and parametric tests that leverage correlation between endpoints for increased power.<sup>48</sup> Gate-keeping procedures address trials with primary and secondary hypothesis families, where secondary analyses proceed only after gatekeeper hypotheses are rejected.<sup>49</sup> Li et al.’s accessible tutorial uses a question-and-answer format organised by the PICOTA framework, addressing common scenarios including multi-arm trials, multiple outcomes, and subgroup analyses.<sup>50</sup> Table 7 summarises appropriate statistical methods for different study designs and outcome types.

Table 7: Statistical methods by study design and outcome type

Study Design	Continuous	Binary	Time-to-Event	Common pitfalls
<b>Cross-sectional</b>	Linear regression	Logistic regression	N/A	Cannot establish temporality; prevalent cases over-represent slowly progressing disease; selection bias if sampling non-random
<b>Cohort</b>	Mixed models	GEE, mixed logistic	Cox regression, frailty models	Loss to follow-up bias; immortal time bias; time-varying confounding unaddressed; informative censoring
<b>Case-control</b>	Conditional regression	Conditional logistic	N/A	Recall bias; inappropriate control selection; over-matching reducing power; cannot estimate incidence
<b>RCT (parallel)</b>	t-test, ANCOVA	Chi-square, logistic	Log-rank test, Cox regression	Ignoring baseline imbalance; per-protocol analysis introducing bias; inter-eye correlation in bilateral studies
<b>RCT (crossover)</b>	Mixed models	GEE	Frailty models	Carryover effects; period effects; dropout before cross-over; inappropriate for conditions that change over time
<b>Platform trial</b>	Bayesian models	Bayesian logistic	Bayesian survival	Prior specification sensitivity; multiplicity from shared control; operational complexity; changing standard of care



**Figure 1:** Bretz graphical approach illustrated for an anti-VEGF dose-finding trial. Panel A shows the initial graph with  $\alpha = 0.05$  distributed equally among four hypotheses. Panel B shows the updated graph after  $H_1$  is rejected, with its  $\alpha$  redistributed via the outgoing edges (weights = 0.5 each) to  $H_2$  and  $H_3$ . VA = visual acuity; CRT = central retinal thickness.

## Reporting standards and quality assurance

Transparent and complete reporting enables critical appraisal and replication. The past five years have seen substantial updates to reporting guidelines reflecting methodological advances. The CONSORT 2025 statement represents the first major update since 2010, integrating advances from recent extensions and addressing contemporary trial features including adaptive designs and estimands.<sup>51</sup> Key guidelines include CONSORT for randomised controlled trials, STROBE for observational studies, TREND for non-randomised evaluations, and STARD for diagnostic accuracy studies.

Adaptive designs require additional reporting elements addressed by the ACE (Adaptive designs CONSORT Extension) statement, a Delphi-developed checklist essential for platform trials.<sup>52</sup> SPIRIT-PRO provides 16 items for including patient-reported outcomes in protocols<sup>53</sup>, while CONSORT-Outcomes 2022 specifies 17 outcome-specific items including the five core elements of outcome definition.<sup>54</sup> These extensions ensure that trials using modern methods are reported with sufficient detail for replication and critical appraisal.

Data sharing transparency became mandatory under ICMJE requirements effective 2018, requiring statements about what data will be shared, when, and through what mechanisms.<sup>55</sup> This represents an ethical obligation arising from trial participants' contributions and enables secondary analyses that can extend the value of trial data. Prospective trial registration (e.g., ClinicalTrials.gov, ISRCTN) prevents selective reporting and publication bias; registration should occur before enrolment begins and include primary outcomes, sample size, and analysis plan.

## Case studies: Applying design principles

### Case 1: SMART design for myopia progression management

**Research question:** What is the optimal sequence of myopia control interventions based on initial treatment response in school-aged children?

**Design:** Sequential Multiple Assignment Randomised Trial with two stages. In Stage 1, children aged 6–12 years with myopia progression  $\geq 0.50$  D/year are randomised to low-dose atropine (0.01%) versus orthokeratology. After 6 months, responders (progression slowed to  $< 0.25$  D/year) continue assigned treatment, while non-responders are re-randomised to either add the alternative treatment or switch to combination therapy from the outset.

**Estimand specification:** The primary estimand uses a treatment policy strategy, comparing embedded dynamic treatment regimens regardless of adherence, answering the question most relevant to clinical practice: ‘If we follow this decision rule, what outcomes can we expect?’

**Key considerations:** Standardised cycloplegic refraction and axial length measurement protocols ensure consistent outcome assessment. The 6-month response criterion balances early identification of non-responders against sufficient time for treatment effects to manifest. Sample size calculations accounting for the sequential design require approximately 300 children to compare the four embedded dynamic treatment regimens with 80% power.

## Case 2: Platform trial for diabetic macular oedema

**Research question:** Which anti-VEGF treatment strategies provide optimal visual outcomes with acceptable treatment burden in diabetic macular oedema?

**Design:** Bayesian adaptive platform trial with shared infrastructure evaluating multiple anti-VEGF agents and dosing regimens. Initial arms include standard monthly aflibercept, treat-and-extend aflibercept, faricimab, and a novel agent entering phase III testing. Response-adaptive randomisation allocates more patients to better-performing arms. Ineffective arms are dropped when posterior probability of inferiority exceeds 95%, while new treatments can be added as they become available.

**Estimand specification:** A composite estimand defines success as gaining  $\geq 10$  ETDRS letters at 12 months without requiring more than 8 injections, capturing both efficacy and treatment burden within the primary endpoint.

**Key considerations:** Central reading centre grading of OCT and visual acuity ensures unbiased outcome assessment. The shared control arm improves efficiency while Bayesian updating enables continuous learning. Governance structures accommodate adding and dropping arms while maintaining trial integrity. Statistical analysis uses hierarchical Bayesian models borrowing information across similar treatment regimens.

## Case 3: Hybrid decentralised trial for dry eye disease

**Research question:** Does a novel anti-inflammatory agent improve symptoms and signs in moderate-to-severe dry eye disease compared to standard artificial tears?

**Design:** Hybrid decentralised double-masked randomised controlled trial. Screening, baseline assessment, and visits at weeks 4 and 12 occur at clinical sites for slit-lamp examination, tear break-up time, and corneal staining. Daily symptom diaries (OSDI subscales) are collected via smartphone application. Weekly quality of life assessments use validated electronic patient-reported outcome measures transmitted directly to the study database.

**Estimand specification:** The primary estimand uses a treatment policy strategy for the symptom endpoint (mean change in OSDI at week 12) and a hypothetical strategy for the sign endpoint (corneal staining), estimating the effect that would have occurred had patients not used rescue artificial tears.

**Key considerations:** The hybrid approach enables more frequent symptom assessment than traditional site-based trials while maintaining rigorous clinical examination at key timepoints. Electronic data capture reduces transcription errors and enables real-time monitoring. Geographic reach expands beyond traditional academic centres while maintaining regulatory-quality data. Validated home visual acuity testing supplements site-based best-corrected visual acuity measurement.

## Discussion

This comprehensive review demonstrates that clinical research methodology in eye and vision research has evolved substantially beyond the traditional parallel-group RCT paradigm. The diversity of approaches now available – from observational methods strengthened by target trial emulation to adaptive platform trials and SMART designs – enables researchers to match study design to specific research questions while maximising both efficiency and scientific rigour.

The unique challenges in eye research require careful consideration throughout the research process. The paired nature of eyes offers opportunities for efficient within-person comparisons but demands appropriate statistical methods to avoid inflated Type I errors. The difficulty of masking many ophthalmic interventions necessitates creative approaches to minimise bias, such as using masked outcome assessors, central reading centres, and standardised protocols. The long follow-up required for progressive diseases like glaucoma creates logistical and financial challenges that registry-based and pragmatic designs may help address.

The integration of decentralised trial elements represents an important evolution, expanding access to research participation while enabling more frequent outcome assessment. However, the specialised examinations fundamental to eye and vision research – including slit-lamp biomicroscopy, tonometry, OCT, and perimetry – remain site-based, making hybrid designs optimal for most ophthalmic research. The systematic review imperative before new trials ensures that research resources are directed toward genuine knowledge gaps, while updated reporting standards including CONSORT 2025 and the ACE statement ensure that methodological

innovations are communicated with sufficient detail for replication and critical appraisal.

Looking forward, several priorities emerge for advancing clinical research in eye care. Enhanced statistical literacy through continued education for clinicians and researchers on appropriate design and analysis methods remains fundamental. Collaborative approaches through multi-centre studies and research networks achieve adequate sample sizes, particularly for rare conditions and for platform infrastructure. Patient-centred outcomes emphasising quality of life and functional outcomes meaningful to patients should guide endpoint selection. Methodological innovation in developing and validating new trial designs suited to ophthalmic conditions will continue advancing the field. Finally, transparency and reproducibility through adherence to reporting guidelines and data sharing practices maximise the value derived from research investments and participant contributions.

## Conclusion

Rigorous clinical research design and analysis form the foundation of evidence-based eye care. This article has provided comprehensive guidance for selecting appropriate study designs, calculating sample sizes, implementing randomisation, choosing endpoints, and applying statistical methods specific to eye and vision research and optometry research. Critically, we have integrated traditional approaches with contemporary innovations including the estimands framework, platform trials, SMART designs, Bayesian adaptive methods, and decentralised trial elements.

The progression from observational studies identifying associations to randomised trials establishing causality represents a continuum of evidence generation. Each design serves important purposes, and the optimal choice depends on the research question, available resources, ethical considerations, and practical constraints. Non-randomised designs strengthened by target trial emulation and propensity score methods offer valuable insights when randomisation is not feasible, while modern experimental designs including adaptive and platform trials provide unprecedented efficiency for comparing multiple treatments.

Statistical considerations permeate every aspect of clinical research, from initial design through final analysis. Proper handling of paired eye data, appropriate methods for missing data, control of multiplicity through graphical approaches, and alignment of analysis with pre-specified estimands are essential for valid conclusions. The choice of endpoints should balance clinical relevance with measurement reliability, and analysis plans should be pre-specified to prevent data-driven decisions that inflate Type I errors.

As we advance toward precision medicine in eye care, the integration of genomic data, imaging biomarkers, and real-world evidence will require sophisticated analytical approaches while maintaining fundamental principles of good research practice. The ethical imperative to review existing evidence before initiating new trials, combined with updated reporting standards ensuring transparency, positions the field

to generate robust evidence that improves patient outcomes. Building on the statistical foundations established in Parts 1 and 2 of this series, clinicians and researchers can contribute to evidence generation that advances our understanding of eye diseases while reducing research waste and accelerating translation to clinical practice.

The next article in this series will explore data visualisation techniques for effectively communicating research findings, while Part 5 will address the ethical dimensions of statistical analysis in eye research. Together, this series aims to enhance statistical literacy across the eye care community, ultimately supporting better research and improved patient care.

## Conflict of Interest Declaration

The author declares that they have no affiliations with or involvement in any organisation or entity with any financial interest in the subject matter or materials discussed in this manuscript.

## Funding Statement

This article did not receive a specific grant from public, commercial, or not-for-profit funding agencies.

## Author Contributions

Daniela Oehring was the principal author and initiator of the article and was responsible for its conception, drafting, and writing.

## Corresponding author



**Prof. Dr. Daniela Oehring**

E-Mail:  
daniela.oehring@plymouth.ac.uk

## References

- Oehring, D., Serra, P. M. (2025). Advancing Statistical Literacy in Eye Care: A Series for Enhanced Clinical Decision-Making Part 1: Introduction to Statistical Tools for Eye Care Research. *Optom. Contact Lenses*, 5, 1, 16–37.
- Oehring, D. (2025). Advancing Statistical Literacy in Eye Care: A Series for Enhanced Clinical Decision-Making Part 2: Statistical Testing in Eye and Vision Research – Principles and Practices. *Optom. Contact Lenses*, 5, 7, 218–237.
- Attebo, K., P. Mitchell, P., Smith, W. (1996). Visual acuity and the causes of visual loss in Australia. The Blue Mountains Eye Study. *Ophthalmology*, 103, 357–364.
- Twelker, J. D., Mitchell, G. L., Messer, D. H., Bhakta, R., Jones, L. A., Mutti, D. O., Cotter, S. A., Klenstein, R. N., Manny, R. E., Zadnik, K.; CLEERE Study Group. (2009). Children's Ocular Components and Age, Gender, and Ethnicity. *Optom. Vis. Sci.*, 86, 918–935.
- Hofman, A., Breteler, M. M., van Duijn, C. M., Krestin, G. P., Pols, H. A., Stricker, B. H., Tiemeier, H., Uitterlinden, A. G., Vingerling, J. R., Witteman, J. C. (2007). The Rotterdam Study: objectives and design update. *Eur. J. Epidemiol.*, 22, 819–829.

- 6 Dong, R., Ying, G. S. (2023). Characteristics of Design and Analysis of Ophthalmic Randomized Controlled Trials: A Review of Ophthalmic Papers 2020–2021. *Ophthalmol. Sci.*, 3, 2, 100266.
- 7 Chalmers, I., Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *Lancet*, 374, 86–89.
- 8 Torgerson T., Evans, S., Johnson, B. S., Vassar, M. (2020). The use of systematic reviews to justify phase III ophthalmology trials: an analysis. *Eye (Lond.)*, 34, 2041–2047.
- 9 Lau, J., Antman, E. M., Jimenez-Silva, J., Kupelnick, B., Mosteller, F., Chalmers, T. C. (1992). Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N. Engl. J. Med.*, 327, 248–254.
- 10 Clarke, M., S. Hopewell, Chalmers, I. (2010). Clinical trials should begin and end with systematic reviews of relevant evidence: 12 years and waiting. *Lancet.*, 376, 20–21.
- 11 Jia, Y., Li, B., Yang, Z., Li, F., Zhao, Z., Wei, C., Yang, X., Jin, Q., Liu, D., Wei, X., Yost, J., Lund, H., Tang, J., Robinson, K. A. (2023). Trends of Randomized Clinical Trials Citing Prior Systematic Reviews, 2007–2021. *JAMA Netw. Open*, 6, e234219.
- 12 Richardson, W. S., Wilson, M. C., Nishikawa, J., Hayward, R. S. (1995). The well-built clinical question: a key to evidence-based decisions. *ACP J. Club.*, 123, A12–13.
- 13 Hulley, S. B., Cummings, S. R., Browner, B. S., Grady D. G., Newman T. B. (2007). *Designing Clinical Research*. Lippincott Williams & Wilkins, Philadelphia.
- 14 Piaggio, G., Elbourne, D. R., Pocock, S. J., Evans, S. J., Altman, D. G.; CONSORT Group. (2012). Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA*, 308, 2594–604.
- 15 Ho, P. M., Peterson, P. N., Masoudi, F. A. (2008). Evaluating the evidence: is there a rigid hierarchy? *Circulation*, 118, 1675–1684.
- 16 Bierman, S. (2023). Flattening the Truth Pyramid: Reconsidering the Evidence Hierarchy. *The Journal of Medical Research*, 9, 157–162.
- 17 Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *American Psychologist*, 39, 334–345.
- 18 Mayo, E. (1933). *The Human Problems of an Industrial Civilization*. Macmillan Company, New York.
- 19 Roethlisberger, F. J., Dickson, W. J. (1939). *Management and the Worker*. 1939, Cambridge, MA, Harvard University Press.
- 20 Hernan, M. A., Robins, J. M. (2016). Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am. J. Epidemiol.*, 183, 758–764.
- 21 Hernan, M. A., Wang, W., Leaf, D. E. (2022). Target Trial Emulation: A Framework for Causal Inference From Observational Data. *JAMA*, 328, 2446–2447.
- 22 Concato, J., Corrigán-Curay, J. (2022). Real-World Evidence - Where Are We Now? *N Engl. J. Med.*, 386, 1680–1682.
- 23 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, U., ICH E9(R1) (2019). Addendum on Estimands and Sensitivity Analysis in Clinical Trials. ICH Harmonised Guideline. International Council for Harmonisation: Geneva, [https://www.gmp-navigator.com/files/guidemgr/E9-R1\\_Step4\\_Guideline\\_2019\\_1203.pdf](https://www.gmp-navigator.com/files/guidemgr/E9-R1_Step4_Guideline_2019_1203.pdf). Referencing: 11December 2025.
- 24 Kahan, B. C., Hindley, J., Edwards, M., Cro, S., Morris, T. P. (2024). The estimands framework: a primer on the ICH E9(R1) addendum. *BMJ*, 384, e076316.
- 25 Ratitch, B., Goel, N., Mallinckrodt, C., Bell, J., Bartlett, J. W., Molenberghs, G., Singh, P., Lipkovich, I., O’Kelly, M. (2020). Defining Efficacy Estimands in Clinical Trials: Examples Illustrating ICH E9(R1) Guidelines. *Ther. Innov. Regul. Sci.*, 54, 370–384.
- 26 Fletcher, C., C. Heffing, N., Wright, M., Bell, J., Anzués-Cabrera, J., Wright, D., Lynggaard, H., Schueler, A. (2022). Marking 2-Years of New Thinking in Clinical Trials: The Estimand Journey. *Ther. Innov. Regul. Sci.*, 56, 637–650.
- 27 RECOVERY Collaborative Group; Horby, P., Lim, W. S., Emberson, J. R., Mafham, M., Bell, J. L., Linsell, L., Staplin, N., Brightling, C., Ustianowski, A., Elmahi, E., Prudon, B., Green, C., Felton, T., Chadwick, D., Rege, K., Fegan, C., Chappell, L. C., Faust, S. N., Jaki, T., Jeffery, K., Montgomery, A., Rowan, K., Juszczak, E., Baillie, J. K., Haynes, R., Landray, M. J. (2021). Dexamethasone in Hospitalized Patients with Covid-19. *N. Engl. J. Med.*, 384, 693–704.
- 28 Woodcock, J., LaVange, L. M. (2017). Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both. *N. Engl. J. Med.*, 377, 62–70.
- 29 Adaptive Platform Trials, Coaliton. (2019). Author Correction: Adaptive platform trials: definition, design, conduct and reporting considerations. *Nat. Rev. Drug Discov.*, 18, 797–807.
- 30 Berry, S. M., Connor, J. T., Lewis, R. J. (2015). The platform trial: an efficient strategy for evaluating multiple treatments. *JAMA*, 313, 1619–1620.
- 31 Park, J. J. H., Siden, E., Zoratti, M. J., Dron, L., Harari, O., Singer, J., Lester, R. T., Thorlund, K., Mills, E. J. (2019). Systematic review of basket trials, umbrella trials, and platform trials: a landscape analysis of master protocols. *Trials*, 20, 572.
- 32 Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. *Stat. Med.*, 24, 1455–1481.
- 33 Almirall, D., Nahum-Shani, I., Sherwood, N. E., Murphy, S. A. (2014). Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. *Transl. Behav. Med.*, 4, 260–274.
- 34 Nahum-Shani, I., Qian, M., Almirall, D., Pelham, W. E., Gnagy, B., Fabiano, G. A., Waxmonsky, J. G., Yu, J., Murphy, S. A. (2012). Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychol. Methods*, 17, 457–477.
- 35 Artman, W. J., Nahum-Shani, I., Wu, T., McKay, J. R., Ertefaie, A. (2020). Power analysis in a SMART design: sample size estimation for determining the best embedded dynamic treatment regime. *Biostatistics*, 21, 432–448.
- 36 Nahum-Shani, I., Almirall, D., Yap, J. R. T., McKay, J. R., Lynch, K. G., Freiheit, E. A., Dziak, J. J. (2020). SMART longitudinal analysis: A tutorial for using repeated outcome measures from SMART studies to compare adaptive interventions. *Psychol. Methods*, 25, 1–29.
- 37 Food and Drug Administration (2010). *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials*. U.S. Food and Drug Administration, Silver Spring, MD.
- 38 Muehleman, N., Zhou, T., Mukherjee, R., Hossain, M. I., Roychoudhury, S., Russek-Cohen, E. A. (2023). A Tutorial on Modern Bayesian Methods in Clinical Trials. *Ther. Innov. Regul. Sci.*, 57, 402–416.
- 39 European Medicines Agency (2022) *Heads of Medicines, Guidance on Decentralised Elements in Clinical Trials*. European Medicines Agency, Amsterdam.
- 40 Medicines and Healthcare products Regulatory Agency (MHRA). (2023). *Decentralised Clinical Trials: Guidance for Sponsors and Investigators*, <https://www.gov.uk/government/collections/medicines-clinical-trials>. Referencing: 0000000
- 41 Park, J., Huh, K. Y., Chung, W. K., Yu, K. S. (2024). The landscape of decentralised clinical trials (DCTs): focusing on the FDA and EMA guidance. *Transl. Clin. Pharmacol.*, 32, 41–51.
- 42 Apostolaros, M., Babaian, D., Corneli, A., Forrest, A., Hamre, G., Hewett, J., Podolsky, L., Popat, V., Randall, P. (2020). Legal, Regulatory, and Practical Issues to Consider When Adopting Decentralized Clinical Trials: Recommendations From the Clinical Trials Transformation Initiative. *Ther. Innov. Regul. Sci.*, 54, 779–787.
- 43 Kruizinga, M. D., Stuurman, F. E., Exadaktylos, V., Doll, R. J., Stephenson, D. T., Groeneveld, G. J., Driessen, G. J. A., Cohen, A. F. (2020). Development of Novel, Value-Based, Digital Endpoints for Clinical Trials: A Structured Approach Toward Fit-for-Purpose Validation. *Pharmacol. Rev.*, 72, 899–909.
- 44 Ford, I., Norrie, N. (2016). Pragmatic Trials. *N. Engl. J. Med.*, 375, 454–463.
- 45 Loudon, K., Treweek, S., Sullivan, F., Donnan, P., Thorpe, K. E., Zwarenstein M. (2015). The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*, 350, h2147.
- 46 Maurer, W., Bretz, F., Posch, M. (2022). Rationale for the update algorithm of the graphical approach to sequentially rejective multiple test procedures. *Pharm. Stat.*, 21, 757–763.
- 47 Bretz, F., Maurer, W., Brannath, W., Posch, M. (2009). A graphical approach to sequentially rejective multiple test procedures. *Stat. Med.*, 28, 586–604.
- 48 Bretz, F., Posch, M., Glimm, E., Klinglmueller, F., Maurer, W., Rohmeyer, K. (2011). Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom. J.*, 53, 894–913.
- 49 Dmitrienko, A., Offen, W. W., Westfall, P. H. (2003). Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Stat. Med.*, 22, 2387–400.
- 50 Li, G., Taljaard, M., Van den Heuvel, E. R., Levine, M. A., Cook, D. J., Wells, G. A., Devereaux, P. J., Thabane, L. (2017). An introduction to multiplicity issues in clinical trials: the what, why, when and how. *Int. J. Epidemiol.*, 46, 746–755.

- 51 Hopewell, S., Chan, A. W., Collins, G. S., Hróbjartsson, A., Moher, D., Schulz, K. F., Tunn, R., Aggarwal, R., Berkwits, M., Berlin, J. A., Bhandari, N., Butcher, N. J., Campbell, M. K., Chidebe, R. C. W., Elbourne, D., Farmer, A., Fergusson, D. A., Golub, R. M., Goodman, S. N., Hoffmann, T. C., Ioannidis, J. P. A., Kahan, B. C., Knowles, R. L., Lamb, S. E., Lewis, S., Loder, E., Offringa, M., Ravaud, P., Richards, D. P., Rockhold, F. W., Schriger, D. L., Siegfried, N. L., Staniszewska, S., Taylor, R. S., Thabane, L., Torgerson, D., Vohra, S., White, I. R., Boutron, I. (2025). CONSORT 2025 statement: updated guideline for reporting randomised trials. *Lancet*, 14, 389, e081123.
- 52 Dimairo, M., Pallmann, P., Wason, J., Todd, S., Jaki, T., Julious, S. A., Mander, A. P., Weir, C. J., Koenig, F., Walton, M. K., Nicholl, J. P., Coates, E., Biggs, K., Hamasaki, T., Proschan, M. A., Scott, J. A., Ando, Y., Hind, D., Altman, D. G.; ACE Consensus Group. (2020). The Adaptive designs CONSORT Extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. *BMJ*, 369, m115.
- 53 Calvert, M., Kyte, D., Mercieca-Bebber, R., Slade, A., Chan, A. W., King, M. T.; the SPIRIT-PRO Group; Hunn, A., Bottomley, A., Regnault, A., Chan, A. W., Ells, C., O'Connor, D., Revicki, D., Patrick, D., Altman, D., Basch, E., Velikova, G., Price, G., Draper, H., Blazeby, J., Scott, J., Coast, J., Norquist, J., Brown, J., Haywood, K., Johnson, L. L., Campbell, L., Frank, L., von Hildebrand, M., Brundage, M., Palmer, M., Kluetz, P., Stephens, R., Golub, R. M., Mitchell, S., Groves, T. (2018). Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. *JAMA*, 319, 483–494.
- 54 Butcher, N. J., Monsour, A., Mew, E. J., Chan, A. W., Moher, D., Mayo-Wilson, E., Terwee, C. B., Chee-A-Tow, A., Baba, A., Gavin, F., Grimshaw, J. M., Kelly, L. E., Saeed, L., Thabane, L., Askie, L., Smith, M., Farid-Kapadia, M., Williamson, P. R., Szatmari, P., Tugwell, P., Golub, R. M., Monga, S., Vohra, S., Marlin, S., Ungar, W. J., Offringa, M. (2022). Guidelines for Reporting Outcomes in Trial Reports: The CONSORT-Outcomes 2022 Extension. *JAMA*, 328, 2252–2264.
- 55 Taichman, D. B., Sahni, P. A., Peiperl, L., Laine, C., James, A., Hong, S. T., Haileamlak, A., Gologly, L., Godlee, F., Frizelle, F. A., Florenzano, F., Drazen, J. M., Bauchner, H., Baethge, C., Backus, J. (2017). Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors. *Ethiop. J. Health Sci.*, 27, 315–318.

## Supplementary File

### Practical Resources

#### Sample Size Calculation Software

G\*Power is a free comprehensive tool supporting sample size calculations for various study designs including t-tests, ANOVA, regression, and chi-square tests. PASS provides commercial software with extensive capabilities for clinical trial designs including adaptive and Bayesian approaches. R packages including *pwr*, *powerAnalysis*, and *simr* enable simulation-based calculations for complex designs. Online calculators such as OpenEpi and Sealed Envelope offer accessible tools for common scenarios.

#### Statistical Analysis Software

R provides free, open-source software with extensive packages for all analyses described in this article, including *lme4* for mixed models, *geepack* for GEE, *survival* for time-to-event analysis, and *rstanarm* for Bayesian methods. SAS remains the industry standard for clinical trials with robust regulatory acceptance. Stata offers user-friendly interfaces with strong epidemiology features. Python with *statsmodels*, *scipy*, and *lifelines* provides flexible programming options for custom analyses.

### Key formulas

Standard error for difference in means:

$$SE = \sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}$$

Sample size for comparing two means:

$n = 2\sigma^2(Z_{1-\alpha/2} + Z_{1-\beta})^2 / \delta^2$ , where  $\sigma$  is the standard deviation, Z values are from the standard normal distribution, and  $\delta$  is the minimum detectable difference.

Design effect for correlated data:

$DE = 1 + (m - 1) \times \rho$ , where  $m$  is the number of observations per cluster and  $\rho$  is the intracluster correlation coefficient.

Non-inferiority margin sample size:

$n = 2\sigma^2(Z_{1-\alpha} + Z_{1-\beta})^2 / (\delta - M)^2$ , where  $M$  is the non-inferiority margin.