DESIGN OF VACCINE EFFICACY TRIALS TO BE USED DURING PUBLIC HEALTH EMERGENCIES – POINTS OF CONSIDERATIONS AND KEY PRINCIPLES

CONTENTS

1 ACKNOWLEDGMENTS

The R&D Blueprint team of the World Health Organization (WHO) gratefully acknowledges the contributions that many individuals made to the development of this guideline. Work was initiated and coordinated by Ana Maria Henao-Restrepo and Pierre-Stéphane Gsell (WHO). The development of this document was led by Doctor Natalie E. Dean and Professor Ira M. Longini, as part of the WHO Working Group 1 for vaccine evaluation, which includes: Ron Brookmeyer, Victor de Gruttola, Christl Donnelly, Elizabeth Halloran, Mamodou Jasseh, Martha Nason, Ximena Riveros, and Conall Watson. WHO extend sincere thanks to the technical experts engaged in the three other WHO Working Groups for vaccine evaluation and who extensively contributed to the development of this document: Steve Bellan, Nele Berthels, Anton Camacho, Peter Dull, John Edmunds, Rosalind Eggo, Susan Ellenberg, Tom Fleming, Lourdes Garcia-Garcia, Julian Higgins, Peter Horby, Philip Krause, Adam Kucharski, Barbara Mahon, Frank Odhiambo, Marie-Pierre Preziosi, Peter Smith, Jonathan Sterne, Alessandro Vespignani, and Yazdan Yazdanpanah.

2 ABBREVIATIONS AND ACRONYMS

AVAREF	African Vaccine Regulatory Framework
BCG	Bacillus Calmette-Guérin
BSL	Biosafety level
CCHF	Crimean-Congo haemorrhagic fever
cRCT	cluster randomized controlled trial
DOI	Declaration of interest
DSMB	data and safety monitoring board
ELISA	Enzyme-linked immunosorbent assay
EMA	European Medicines Agency
EUA	Emergency Use Authorization
EUAL	Emergency use assessment and listing
EVD	Ebola virus disease
FDA	Food and Drug Administration
GCP	Good Clinical Practice
GEE	Generalized estimating equation
GMP	Good Manufacturing Practice
HCW	Healthcare workers
HIV	Human immunodeficiency virus
HPTN	HIV Prevention Trials Network
ICC	Intracluster (intraclass) correlation coefficient
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IRB	Institutional review board
iRCT	individually randomized controlled trial
ITT	Intention-to-treat
IU	International unit
MERS-CoV	Middle East respiratory syndrome coronavirus
mITT	modified intention-to-treat
NAM	National Academy of Medicine
NI	Non-inferiority

NRA	National regulatory authority
PCR	Polymerase chain reaction
PHE	Public health emergency
PHEIC	Public health emergency of international concern
PP	Per-protocol
PRNT	Plaque reduction neutralization test
R&D	Research & development
R0	Basic reproduction number
RDT	Rapid diagnostic test
RVF	Rift Valley Fever
SAP	Statistical analysis plan
SARS	Severe acute respiratory syndrome
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TPP	Target product profile
UN	United Nations
VE	Vaccine efficacy or vaccine effectiveness
WHO	World Health Organization

3 GLOSSARY OF TERMS

- *Correlate of risk*: a biomarker that predicts efficacy
- *Disease X*: any future disease that meets the Blueprint criteria for a priority disease
- *Indirect vaccine effectiveness:* effect of the vaccine in reducing infection/disease in unvaccinated individuals who are in the same populations as vaccinated individuals
- *Overall vaccine effectiveness:* population-level effect of the vaccine in reducing infection/disease, where the population may include both vaccinated and unvaccinated individuals
- *Priority disease*: disease selected by a committee of experts because of its epidemic potential and lack of adequate medical countermeasures
- *Public Health Emergency of International Concern*: WHO classification for an extraordinary event that constitutes a public health risk to other countries through the international spread of disease, and potentially requires a coordinated international response
- *Responsive trial*: trial design in which trial sites or trial participants are identified in response to local transmission of the target pathogen
- *Surrogate of protection*: a validated replacement biomarker for efficacy
- *Total vaccine effectiveness*: the combined effect of direct vaccine protection and indirect vaccine effectiveness
- *Vaccine efficacy*: an individual-level measure of the effect of the vaccine in reducing infection/disease among vaccinated persons

4 EXECUTIVE SUMMARY

Under the Research & Development (R&D) Blueprint Plan of Action, the World Health Organization (WHO) has convened a group of experts in clinical trials, regulatory, and outbreak management, to agree on standard procedures to rapidly evaluate experimental vaccines during public health emergencies (PHEs) while maintaining the highest scientific and ethical standards. This guidance document details major vaccine study designs that could be used during outbreaks and PHEs of emerging and re-emerging pathogens for which there is no licensed vaccine.

The development of this document was framed by the WHO Blueprint diseases, prioritized for their likelihood to cause PHEs and the lack of available medical countermeasures. The unusual nature of PHEs caused by emerging pathogens (e.g., lack of knowledge on disease natural history and transmission, lack of characterization of the experimental vaccine, resource-limited settings, unpredictable duration and location of outbreak) creates unique conditions which pose methodological and practical challenges to vaccine evaluation in affected countries. Lessons learned from past outbreaks underscore the need for innovative and flexible vaccine study designs and analysis in order to provide valid and robust scientific evidence on the performance of an experimental vaccine at the earliest opportunity.

We describe the regulatory principles underlying the process of vaccine evaluation before detailing the major design planning considerations, including definition of study endpoints and populations. We outline available study designs that may be used during PHEs. For each study design, we provide general descriptions, examples, and suggestions for how to customize the design for outbreaks of emerging pathogens. We introduce both randomized trial designs and observational study designs, though we strongly recommend randomized trials with rare specified exceptions.

Trial design strategies should be tailored to the context of the disease, the particular outbreak, the local setting, and the vaccine candidate to be evaluated. Though trials in the general population are generalizable to a broader population, high risk populations may be targeted to maximize the power of the trial. In settings where incidence is spatiotemporally unpredictable, responsive strategies in which site/participant enrolment and vaccination is triggered by local transmission may be preferred. These strategies work best when there are minimal delays in testing and data collection, with fast-acting vaccines, and where research infrastructure and regulatory structures allow for rapid approval and initiation.

Individual randomization is generally preferred over cluster randomization because it is not necessary to inflate the sample size to account for clustering. Cluster randomized trials may be preferred in settings where individual randomization is not accepted by the trial population or where estimation of indirect and overall vaccine effects is prioritized. Stepped wedge trials are challenging to plan, implement, and interpret in the context of PHEs, and they can lack valuable flexibility. Multi-arm and factorial trials may be considered to simultaneously evaluate multiple interventions. It is important for studies to plan for settings where the epidemic may decline. Studies may consider combining information across outbreaks to accumulate more evidence to estimate efficacy.

This guidance outlines generic principles and methodological elements on how to best design, implement and analyse vaccine trials during PHEs. For each priority pathogen, WHO will convene experts to apply and tailor the proposed methodology to generate a pathogen-specific vaccine

evaluation guidance. Advance planning for vaccine trial designs is critical for rapid and effective response to a PHE and to advance knowledge to address and mitigate future PHEs.

5 Scope and Objectives

5.1 Scope of the Document

This document introduces standard procedures to rapidly evaluate candidate vaccines against the Blueprint priority diseases while maintaining the highest scientific and ethical standards, should candidate vaccines be prioritized for evaluation in countries affected by the Blueprint priority diseases. The Blueprint priority diseases have been selected for their likelihood to cause a PHE and because of the lack of other public health countermeasures available for these diseases, including licensed diagnostics, therapeutics, or vaccines. Typically these pathogens result in outbreaks with unpredictable spatiotemporal incidence, though transmission of some of these pathogens at ongoing low levels may not be classified as a PHE of international concern (PHEIC). This list of priority diseases is being re-evaluated annually by a panel of scientists and public health experts. As of February 2018, the priority diseases are:

- Lassa fever
- Crimean-Congo Haemorrhagic Fever (CCHF)
- Ebola virus disease and Marburg virus disease
- Middle East respiratory syndrome coronavirus (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS)
- Nipah and henipaviral diseases
- Rift Valley Fever (RVF)
- Zika
- Disease X

Disease X represents the knowledge that a serious international epidemic could be caused by a pathogen currently unknown to cause human disease, and so the R&D Blueprint explicitly seeks to enable cross-cutting R&D preparedness that is also relevant for an unknown “Disease X” as much as possible.

An updated list of priority diseases and the methodology used for disease prioritization can be accessed at the following link:

http://www.who.int/blueprint/priority-diseases/en/

This document is designed to outline considerations on how to evaluate preventive vaccine efficacy for vaccine candidates targeted against the Blueprint priority diseases and other emerging pathogens for which no licensed vaccine is available. To a lesser extent, we also consider the setting in which a licensed vaccine exists, but we want to evaluate another vaccine, maybe for superior efficacy or safety. The main scope considered in this document corresponds to the scope of research preparedness as defined in the WHO Blueprint Plan of Action (WHO 2016a).

5.2 Target Audience

The target audience for this document is groups seeking to design studies of vaccine efficacy for emerging pathogens, as well as the national regulatory authorities and ethics boards who will evaluate these studies. The target audience includes the technical community of clinical researchers, whether from academia, industry, or government, who will assist in the design, analysis, and interpretation of vaccine efficacy studies, and who develop methodology that may be implemented in future studies.

We aim to share these recommendations with the wider community for feedback so that they can optimally inform the design and implementation of trials meeting ethical and regulatory approval and with maximal likelihood of success and adequate power to evaluate the candidate vaccine.

5.3 Related Documents

WHO. *An R&D Blueprint for Action to Prevent Epidemics: Plan of Action, May 2016.* http://www.who.int/entity/blueprint/about/r_d_blueprint_plan_of_action.pdf?ua=1
WHO R&D Blueprint. *2017 Annual review of diseases prioritized under the Research and Development Blueprint*. http://www.who.int/blueprint/priority-diseases/en/
WHO R&D Blueprint AP2. *Intervax Tool: Decision Support for Pathogen Outbreak*. http://situatedlaboratories.net/WHO-interactive-flow/dist/
WHO. *Good participatory practice guidelines for trials of emerging (and re-emerging) pathogens that are likely to cause severe outbreaks in the near future and for which few or no medical countermeasures exist (GPP-EP)*. http://www.who.int/blueprint/what/norms-standards/GPP-EPP-December2016.pdf?ua=1
WHO Expert Committee on Biological Standardization. *Guidelines on clinical evaluation of vaccines: regulatory expectations*. http://www.who.int/biologicals/vaccines/clinical_evaluation/en/
WHO R&D Blueprint. *Roadmaps & Target product profiles for the prioritized diseases*. http://www.who.int/blueprint/what/research-development/en/
International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). *ICH Guidelines*. http://www.ich.org/products/guidelines.html

6 BACKGROUND

6.1 GOAL OF VACCINE EFFICACY TRIALS

Candidate vaccines typically proceed through several phases of testing during development. Although the distinctions between trials at different phases of development may sometimes be blurry, trials initially evaluate safety and immunogenicity in a small population (Phase 1) (typically 30-100 healthy human volunteers, often study different doses and/or vaccination schedules, primary focus on safety and tolerability, also designed to provide preliminary assessments of immunogenicity), expand to a larger and more targeted population (Phase 2) in order to obtain safety and immunogenicity data to support a larger subsequent efficacy study, and may culminate in a randomized clinical trial measuring preventive vaccine protection against the disease of interest (Phase 3).

The demonstration of vaccine efficacy (VE), typically through a Phase 3 trial, is fundamental for licensure and to help inform policy-makers about potential uses of vaccines. Vaccine efficacy is an individual-level measure of vaccine effects defined as the proportionate reduction of the incidence of the target infection in vaccinated participants compared to controls (Halloran et al. 1991); generally, vaccine efficacy equals one minus the hazard, odds, or risk ratio, with 100% efficacy corresponding to zero incidence in vaccinated persons. The demonstration of efficacy is especially important for a first vaccine to be licensed against a specific pathogen, as would be the case for the Blueprint priority diseases. Satisfactory evidence of vaccine efficacy and vaccine safety are the primary factors evaluated when determining vaccine licensure (Russek-Cohen et al. 2016).

6.2 CONTEXT OF PUBLIC HEALTH EMERGENCIES

Assessing vaccine efficacy during a PHE of an emerging pathogen presents unique challenges. One of the greatest challenges is spatiotemporal variability in incidence. Cases may be spatially dispersed, with surges in infection in particular areas at different times. It may be very difficult to predict the incidence and location of cases, and accuracy of predictions may be limited by the availability of high-quality surveillance data. As a result, it may be difficult to predefine a target population that has sufficiently high transmission risk to observe the required number of events to assess vaccine efficacy.

Unlike endemic diseases, outbreaks end or are contained to a point such that only sporadic cases continue to occur. As a result, once a trial site is selected for a vaccine for an emerging pathogen, the clinical evaluation process must start at the earliest opportunity in order to capture transmission within the vaccine study. This requires the trial to be planned in advance, to receive expedited protocol approval, and for staff to be rapidly trained. It is also important to quickly evaluate the vaccine because, if the vaccine is efficacious, we want the ability to use the vaccine in that outbreak. If a vaccine appears to provide sufficient public health benefit following its evaluation, it may be used and/or recommended for use under Expanded Access before licensure (Gsell et al. 2017). If the vaccine is highly efficacious, a vaccine trial could contribute to the waning of disease, which would be advantageous for disease control but not necessarily for clinical evaluation. Once the outbreak is

contained, the vaccine likely cannot be evaluated for efficacy in the standard way, and it may be unclear when another opportunity for evaluation would arise.

The optimal study design/protocol may depend on information not immediately available during the planning phase (Piantadosi 2005:2.3.6). If the pathogen has not caused large outbreaks in humans before, we may not be able to accurately estimate the clinical attack rate of the disease (proportion infected and manifesting symptoms), the incubation period (time from infection to symptom onset), mode(s) of transmission, and other basic characteristics of the disease. The required sample size may be impossible to judge accurately without preliminary trials (Piantadosi 2005:2.3.6). If Phase 1 and 2 testing have not been conducted or are still ongoing, we will have less available information about the vaccine and characteristics of the immune response to guide study design. If multiple candidates are being considered, it is possible that we will not know which candidate(s) will be selected for Phase 3 testing. There may be poor diagnostics or diagnostics that are not fully validated or commercially available. These information gaps make it more challenging to design a study optimally tailored to the context.

Other key challenges relate to the logistics of running a trial in a resource-limited setting during a public health emergency. Outbreaks of emerging pathogens tend to occur in developing countries. These countries may have limited medical or laboratory infrastructure (e.g., diagnostics, cold chain) to support a trial, and this infrastructure may be strained by the ongoing outbreak. Medical and technical teams need to be trained in Good Clinical Practice (GCP). Resources may be especially scarce early on in the outbreak, before an international response has been mobilized. There could be limited supply of vaccine, or vaccine may be produced in batches that will become available over time. The available number of doses could impact the sample size and preferred trial design.

There may be fear in the population of the "unknown" pathogen, affecting acceptability of diagnostic testing, contact tracing, vaccination, and other control efforts. Therefore an appropriate community engagement strategy is required in order to design a study acceptable to the affected communities. If the outbreak has caused a humanitarian crisis, there may be mass migration or other forms of instability in the population. There may be key political and social pressures, especially as related to the ethics of the trial design. There may be heightened risk of transmission of pathogens during health care delivery, which may affect risk/benefit ratio for specimen collection, such as drawing blood.

Finally, such trials require immense international coordination. Implementation of a trial would require collaboration of different international organizations, including multiple ethical approvals. For example, the Ebola ça Suffit trial engaged more than 15 different institutions. There may also be other concurrent trials in the same populations, evaluating prevention strategies, treatments, and other interventions. There are expected to be other outbreak response measures, implemented by WHO and other partners. Public health efforts, including careful contact tracing and safe care for infected individuals, must not be impeded by the trial (Lane, Marston, and Fauci 2016).

6.3 Rationale for the Points of Considerations and Guidance

The 2013-2016 West-African Ebola outbreak has broken new ground in speed and study design has a set a precedent for vaccine evaluation during PHE. Based on lessons learned from that outbreak, the WHO's Blueprint Plan of Action aims to reduce the time lag between the declaration of a PHE and the availability of medical countermeasures, including vaccines. One area includes advancing early phase development of vaccine candidates for priority diseases during inter-epidemic periods. Supporting expansion of capacity to implement adequate study design is another area of the Blueprint Plan, which is the motivation for this work. We aim to encourage consensus on properly designed studies by multilateral efforts and provide advance insight into expert opinion on critical issues expected to arise in the future.

We present considerations for the design and analysis of studies for vaccine efficacy, focusing primarily on randomized trials. The Blueprint priority diseases, selected for their likelihood to cause PHEs and the lack of adequate medical countermeasures, are used to frame our methodological discussions. In **Section 7,** we detail the regulatory considerations that guide the evaluation of vaccine efficacy and safety. In **Section 8,** we examine the major design planning considerations for vaccine studies, including defining study endpoints, populations, and comparator arms. In **Section 9**, we review the major study designs to evaluate an experimental vaccine, including randomized trial designs in **Section 9.1** and observational study designs in **Section 9.2**. In **Section 10**, we discuss additional design and conduct considerations, such as data monitoring and missingness. In **Section 11**, we describe operational and implementation issues relevant to studies of emerging pathogens. In **Section 12**, we detail how these recommendations are impacted by characteristics of the disease and the vaccine candidate(s). Finally, in **Section 13**, we describe how these recommendations are to be disseminated and updated in the future.

This document is not a substitute for the expert advice of epidemiologists, statisticians, ethicists, clinicians, and others involved in the design and implementation of vaccine studies. Rather, this work is intended to guide discussion with these experts and, in turn, enhance the quality and utility of study data.

6.4 Declaration of Interests

According to WHO regulations, all external experts must declare their relevant interests prior to participation in the WHO guideline development process and meetings. All members of the WHO Blueprint Working Groups for vaccine evaluation were therefore required to complete a standard WHO Declaration of Interest (DOI) form before engaging in the guidance development process and before participating in related meetings. The WHO R&D Blueprint Secretariat reviewed all DOI forms before finalizing the invitations for experts to participate in the development of the guidance document. None of the contributors have any competing interests.

7 Regulatory Considerations

7.1 Regulatory Principles

The major goal of conducting clinical efficacy studies during PHEs is to obtain data that can support broader use of a vaccine under a defined regulatory framework. It is recognized that some of the relevant data may be collected either before or after the PHE, but it is presumed that efficacy data from a study conducted at a time when disease transmission is ongoing is central for regulatory approval. Because of the confidence that product licensure provides in safety and effectiveness, licensure, with or with a requirement for post-licensure confirmation of product effectiveness, is normally the ultimate goal. Thus, clinical studies are typically aimed at accumulating data that will support this end result.

These data could also support other uses of the vaccine in the interim between trial completion and licensure, including temporary approval through "expanded access" use, Emergency Use Authorization (EUA), or a similar mechanism. This special procedure gives time-limited approval to a product for a disease in a PHEIC. Use of the vaccine must have oversight by a functional national regulatory authority (NRA) in conjunction with the WHO Emergency use assessment and listing (EUAL) for vaccines. EUA/EUAL normally does not require consent, though consent requirements should be considered in conjunction with the data that will be submitted to/reviewed by regulators. EUA determination reflects a lower level of certainty about the effectiveness of the product as compared to licensure. While the degree to which product safety and effectiveness has been established may differ among these various regulatory mechanisms, they all share the need for scientifically valid data that supports product effectiveness. It should be recognized that widespread distribution of a vaccine prior to obtaining sufficient data to support licensure will likely interfere with collecting that data, and may lead to a situation in which licensure is significantly delayed.

During the regulatory decision-making process, clinical trial results will be evaluated by regulators based on principles not only of good clinical practice, but also of good trial design. This includes evaluation of potential biases, speaking to the importance of measures like randomization and blinding that can reduce potential biases in clinical trial outcomes. Previous consultations have concluded that it may be ethical to conduct randomized trials under circumstances when there are insufficient data to support widespread use or licensure of a product (see **Section 9.1.1**). Selection of the primary endpoint for vaccine efficacy trials in PHEs is usually guided by the desired indication for use, taking into account considerations of study power. Usually, this endpoint is the incidence of pathogen-associated disease relative to that of a comparator group (see **Section 8.2.1**). Pre-specified diagnostic criteria, including qualification or validation of assays to be used, are required.

Vaccine studies should have prospectively defined statistical analysis plans, including unambiguous descriptions of the groups to be compared, the specific analyses to be performed, procedures for controlling Type I error (for primary and critical secondary endpoints, as well as both for interim and final analyses), and the specific criteria for success (i.e., lower bound on the confidence interval of the efficacy estimate)**.** Efficacy studies for vaccines may range in size depending on this criteria for success, which should be set based on agreement with regulatory authorities.[1] Considerations in

[1] In a setting of having two trials for licensure, a frequent practice is to set the type 1 (alpha) error for each trial at 0.05. This would correspond to each trial having a (one-sided) false positive error rate of 0.025, which is a rate of reaching a positive conclusion in 1 in 40 trials when the null hypothesis truly holds. A single trial may be sufficient to support licensure if it is a well-controlled/well-designed efficacy study and the estimate of VE is sufficiently high and sufficiently precise that the null hypothesis is rejected at the 'strength of two trials', i.e.,

setting success criteria may include product safety and benefit, the public health need (including incidence and severity of the disease), the availability of other products, and the totality of available data on the product. Sample size calculations are normally also based on power calculations that take into account the expected disease incidence among participants in the study.

Because clinical trials performed during outbreaks may need to adapt to changing epidemiology, any adaptive features of the trial design (including the potential risk of not accruing the planned number of cases) should be prospectively described in the context of the statistical analysis plan (see **Section 10.1.1**). Changing epidemiology may make it difficult to obtain sufficient clinical disease data to directly support vaccine efficacy. For this reason, it can be important to collect as much serological data as possible in these trials (ideally both at baseline and post-vaccination). Such data can potentially support licensure based on "surrogate endpoints" that may be used in alternative pathways such as "Accelerated Approval" in the United States, or "Contingent Approval" in the European Union. These alternative pathways generally require the cumulative data to show a reasonable likelihood that the vaccine is effective prior to licensure, with the requirement to provide additional confirmatory studies after licensure. Scientific considerations for the use of immunological data to support vaccine efficacy are discussed below (see **Section 7.2**). Immunogenicity data in humans can also sometimes be used as a "bridge" to support licensure under the "Animal Rule", a pathway available in the U.S. in which animal efficacy data can be used in lieu of human efficacy data when other means of obtaining efficacy data are not feasible. Typically these alternative strategies require conditional post-licensure (Phase 4) surveillance to obtain additional safety and effectiveness study (Hudgens, Gilbert, and Self 2004; Russek-Cohen et al. 2016).

Although we focus on the collection of vaccine efficacy data, it is important to note that sufficient vaccine safety must be collected at each stage of development to support entry into the subsequent stage (see **Section 8.2.4**). Likewise, each regulatory decision (e.g., licensure, EUA, Expanded Access, etc.) is associated with requirements for sufficient safety data to permit weighing of benefits and risks in the context of these decisions. Formal evaluation of efficacy against serious and rare complications of infection may not always be feasible. Post-marketing safety studies can provide additional important information about the safety of a vaccine.

It is important to achieve agreement among the regulators responsible for making decisions about the use of the product that the study, if its success criteria are met, will support the desired outcome of licensure or otherwise. This may include both regulators in more highly developed countries and regulators in countries where the outbreak is occurring, where regulatory agencies may not be as

---

at the square of 1 in 40, corresponding approximately to having a (one-sided) false positive error rate of 0.0005. For example, if the null hypothesis is VE=30% a single trial may be sufficient to support licensure if the null hypothesis of VE=30% is ruled out not only by a 95% two-sided confidence interval, but also by a 99.9% two-sided confidence interval. In PHEs, it could be argued that a weaker 'single trial' standard might be acceptable, such as requiring in the previous example that the null hypothesis of VE=30% simply be ruled out by a 95% two-sided confidence interval. The acceptability of such an argument should be influenced by many other issues such as the clinical importance of the effect on the efficacy measure, the evidence regarding safety, and the scientific quality of trial conduct. It could be argued that weaker results could potentially support Emergency Use Authorization-type decisions in emergency situations, but it would be dangerous to have different statistical criteria for licensure, and using a weaker criteria in emergency situations may make subsequent licensure more difficult.

well-resourced. Regulatory consensus can sometimes be facilitated by joint review, as well as by adherence to key clinical trial design and regulatory principles such as those spelled out in this and other relevant WHO documents (WHO Expert Committee on Biological Standardization 2017a). Continued discussion with regulators can help to reduce the possibility that changing circumstances could negatively impact the utility of the study.

7.2 Correlates of Risk and Surrogates of Protection

Licensure decisions may be based on immunological data if a valid surrogate of protection exists. Correlates of risk are biomarkers that *correlate* with the clinical endpoint measuring vaccine efficacy. In contrast, validated surrogates of protection are biomarkers that *predict* efficacy; i.e., the intervention's effect on the biomarker provides reliable evidence about the intervention's net effect on vaccine efficacy (Qin et al. 2007). It is recognized that some important biomarkers may fall in the spectrum between these definitions, as regulatory decisions such as U.S. Accelerated Approval may be based on a marker that is reasonably likely to predict protection. It is important to note that validated surrogates of protection do not exist for any of the Blueprint priority diseases.

Validated surrogates of protection could be important in that they could guide the development of vaccines, they could enable reducing the size and duration of definitive trials, and they could support the bridging of a vaccine observed in a trial to new setting. A validated surrogate allows for significantly more efficient evaluation of newer vaccines (such as second generation products) since vaccine efficacy could be indirectly demonstrated through the surrogate, although a large safety study may still be necessary (Hudgens et al. 2004).

To justify the use of biomarkers, such as immune markers, as replacement endpoints, several issues need to be addressed (Fleming and Powers 2012). Is the biomarker in a direct causal pathway of the disease process that is meaningfully impacted by the intervention? Are there other principal causal pathways not captured by the biomarker? Is it known what magnitude and duration of effect on the biomarker is needed to have clinically meaningful effects on clinical endpoints such as vaccine efficacy? Most importantly, does the intervention have off target effects or safety risks that are not captured by the biomarker and yet that could meaningfully impact the clinical endpoint?

Since a sufficiently comprehensive understanding about the causal pathways of the disease process and about the intended and unintended effects of the intervention is rarely achievable, it also is important to have a collection of clinical trials in sufficiently related settings that provide direct insights about the relationship between the intervention's net effect on the biomarker and its net effect on the clinical endpoint. Hence, when feasible, potential biomarkers should be measured in pre- and post-vaccination serum samples to enable their evaluation as correlates of protection.

The Institute of Medicine provided an in-depth consideration of biomarkers as surrogate endpoints, (Institute of Medicine 2010). The Institute of Medicine noted that context of use impacts the validity of a biomarker as a replacement endpoint, in that a validation for one class of interventions may not apply to another. In the setting of vaccines, if immune markers are specific to the vaccine platform, then they may not be applicable to other vaccines for the same disease. For example, for Ebola, the immune markers appear to differ across vaccine platforms. The WHO has also prepared guidance on correlates of vaccine-induced protection (World Health Organization 2013a). For additional technical information on correlates and surrogates of vaccine-induced protection, the following references are

provided (Gabriel, Daniels, and Halloran 2016; Gilbert and Hudgens 2008; M E Halloran, Longini Jr., and Struchiner 2010; Hudgens et al. 2004; Plotkin and Gilbert 2012).

8 Major Design Planning Considerations

The goal of vaccine studies is to estimate vaccine efficacy and/or key vaccine effects, which we review in **Section 8.1**. We then describe potential study endpoints, including how vaccine efficacy will be measured, as well as secondary and safety endpoints (see **Section 8.2**), the study population (see **Section 8.3**), and the study comparator arm (see **Section 8.4**).

8.1 Types of Vaccine Effects

To guide our discussion of vaccine design planning considerations, we first review the different types of vaccine effects estimable from vaccine studies. A detailed overview of the scientific and public health value of each of these vaccine effects is provided in Halloran et al. (2010:2).

When we refer to vaccine efficacy, we are referring to an individual-level measure of vaccine effects, typically direct vaccine effectiveness or, less commonly, total vaccine effects. Direct vaccine effectiveness measures the direct protection conferred to a vaccinated individual. This quantity is estimable from individually randomized trials by comparing rates in vaccinated and unvaccinated individuals in the same populations during the same timeframe.

Indirect vaccine effectiveness measures protection conferred to individuals via the vaccination of others in the same population. They are also known as spillover effects and directly link to the concept of herd immunity. Indirect effects are not directly estimable from individually randomized controlled trials, but they can generally be inferred in cluster randomized controlled trials by comparing unvaccinated individuals in vaccinated populations with unvaccinated individuals in vaccinated populations. Indirect vaccine effectiveness is a function both of the vaccination coverage level in the population and of the ability of the vaccine to reduce disease incidence.

Total vaccine effectiveness combines direct and indirect effects. It can be estimated by comparing incidence in vaccinated individuals in vaccinated populations with unvaccinated individuals in unvaccinated populations. It is the primary measure of VE from cluster randomized controlled trials as vaccinated individuals in vaccine clusters receive both direct protection from vaccine and indirect protection from other vaccinated persons. Total effects are expected to always be larger than direct effects because of the role of indirect effects. Similar to indirect effects, total effects are a function of the vaccination coverage level in the population as well as of the direct effects of the vaccine.

Finally, overall vaccine effectiveness compares the cluster-level incidence in vaccinated and control populations, including all individuals who did not receive the vaccine for any reason. Overall vaccine effectiveness is a valuable public health measure reflecting the overall utility of the strategy. It is a function of the vaccination coverage level in the population as well as of the direct effects of the vaccine.

8.2 STUDY ENDPOINTS

For a given pathogen, study endpoints should be selected to support the broader intended use of a vaccine. To enhance interpretability, typically a single primary endpoint and up to 3 or 4 secondary endpoints are defined in the study protocol. Co-primary endpoints, if used, require some adjustment for multiple testing so that there is a clear criteria for success that controls the type I error rate of the study.

8.2.1 PRIMARY ENDPOINT

The ideal primary endpoint in a Phase 3 trial should directly measure the disease-related outcome of public health interest, such as survival or disease-related symptoms, which supports the vaccine indication for use. Disease-specific characteristics as extracted from a selection of pathogen's WHO vaccine target product profile (TPP) are summarized in Table **1**. In vaccine trials, the most common primary endpoint is clinical disease with laboratory confirmation, though there are some settings where other primary endpoints are selected for practical purposes. For example, the desired endpoint may be rare, leading to an infeasible sample size. If diagnostics are poor and/or infrastructure is limited, endpoints requiring laboratory confirmation may be hard to detect. A replacement endpoint, such as a biomarker, may be used where there is substantive evidence validating that the effect of the vaccine on the replacement endpoint reliably predicts the vaccine's effect on the disease-related outcome of interest. We describe potential choices for primary endpoint and the settings in which they may be preferred. A further review of endpoints in vaccine trials is available in Hudgens et al. (2004).

	Indication for Use	**Preferred measure of efficacy**	**Preferred target population**
Ebola	For immunization of at-risk persons residing in the area of an on-going outbreak to protect against Ebola virus disease caused by circulating species of filovirus; to be used in conjunction with other control measures to curtail or end an outbreak	Preventing disease in healthy adults, adolescents and children	All age-groups and populations at high present risk of Ebola virus disease caused by circulating species of filovirus
Lassa	For active immunization of persons considered potentially at-risk, based on specific risk factors, to protect against Lassa disease.	Preventing infection or disease	All age groups. Suitable for administration to pregnant women.

MERS-CoV	For active immunization of persons considered at-risk based on specific risk factors to protect against MERS-CoV. Risk groups will include health care workers, frontline workers and those working with potentially infected animals.	Preventing Middle East Respiratory Syndrome caused by MERS-CoV in healthy adults	Health care workers, frontline workers and others with occupational risk. Suitable for administration to pregnant women.
Nipah	For active immunization of at-risk persons in the area of an on-going outbreak for the prevention of Nipah disease; to be used in conjunction with other control measures to curtail or end an outbreak	Preventing Nipah infection or disease in healthy adults	All age groups and populations at high risk of Nipah disease
Zika	For the prevention of Zika virus-associated clinical illness of any severity in subjects 9 years of age or older	Prevention of virologically-confirmed Zika illness	Women of reproductive age (including adolescent and pre-adolescent girls 9 years of age or older), and boys/men of the same ages.

TABLE 1 – EXAMPLE OF PREFERRED CHARACTERISTICS EXTRACTED FROM THE WHO VACCINE TARGET PRODUCT PROFILES (ANON N.D.) FOR EBOLA, LASSA, MERS-COV, NIPAH AND ZIKA, SOME OF THE PATHOGENS PRIORITIZED BY WHO.

8.2.1.1 Clinical Disease with Laboratory Confirmation

The most common primary endpoint is clinical (symptomatic) disease with laboratory confirmation. Generally public health interest is in preventing or lessening disease rather than preventing infection (Clements-Mann 1998). There are examples of vaccines that have been shown to prevent disease but not infection, including rubella, mumps, measles, and polio (Clements-Mann 1998).

With this kind of endpoint, laboratory assays may confirm infection (by polymerase chain reaction (PCR) or antigen-detection assays) or confirm seroconversion when symptoms are present. A baseline/pre-vaccination sample would likely be necessary to assess seroconversion. If the pathogen is detectable in the host for only a limited period of time, pathogen-detection assays may have reduced sensitivity resulting in false negative endpoints. Cases may also be undercounted if poor laboratory infrastructure results in many missing or failed confirmatory tests (Nason 2016). On the other hand, failure to use an adequately specific test to confirm the pathogen can result in many false positives. Laboratory confirmation is particularly important when the disease symptoms are not specific enough to rule out other aetiologies.

8.2.1.2 Clinical Disease without Laboratory Confirmation

Another potential choice for primary endpoint is clinical disease without requiring laboratory confirmation. Such an endpoint may be considered in settings where laboratory confirmation of all cases is extremely difficult due to poor diagnostics or limited local infrastructure.

This type of endpoint has key limitations for pathogens without a highly distinct clinical syndrome. A case definition with poor specificity decreases estimates of vaccine efficacy, as has been shown for rotavirus vaccination (Lachenbruch 1998). This attenuation is most apparent in settings where

diseases with similar symptoms are common, especially when the target disease is rare. Studies using clinical disease without laboratory confirmation as the primary endpoint should consider validation sets, which are random samples of cases (and possibly a smaller subset of controls) that are tested for laboratory confirmation (Halloran and Longini 2001). Internal estimates of sensitivity and specificity of the case definition can be used to derive adjusted estimates of vaccine efficacy (Lachenbruch 1998).

8.2.1.3 Infection with Laboratory Confirmation

Limited settings exist where infection, with or without clinical disease, would be the preferred endpoint. As infection alone is rarely the outcome of public health interest, using it as the primary endpoint in a vaccine efficacy study would be primarily motivated by practical considerations.

Infection may be useful when it serves as a replacement endpoint for another endpoint that is rare or difficult to measure. For example, Zika congenital syndrome is an important outcome of public health interest, but it is too rare to be used as the primary endpoint in an adequately powered trial. Since Zika congenital syndrome may result from either symptomatic or asymptomatic infection in pregnant women, infection may be a useful replacement endpoint. Infection endpoints would also work well for diseases with long incubation periods, such as HIV and tuberculosis, in which a traditional endpoint of clinical disease would not be feasible given the length of the trial (Hudgens et al. 2004).

It may also be difficult to power a study for a clinical disease endpoint for pathogens with low pathogenicity (high asymptomatic rate). An infection endpoint may make the trial more feasible in terms of required sample size by increasing the event rate. An important limitation of this approach is that it would fail to capture the effect of a vaccine that works by reducing disease severity rather than blocking infection. Furthermore, it could be possible for a vaccine to reduce infection but have little impact on the disease of public health interest, especially for pathogens with multiple strains or serotypes. This was a concern in the design of human papillomavirus (HPV) trials, where the public health outcome of cervical or other anogenital cancers was not a feasible endpoint. Though the largest HPV trials used a symptomatic pre-cancer efficacy endpoint, at least one trial used HPV 16/18 infection persisting for at least one year as the primary endpoint (Schiller, Castellsagué, and Garland 2012).

The feasibility of an infection endpoint will depend on the quality of available diagnostics. Seroconversion could be detected using baseline (pre-vaccination) and follow-up samples where a reliable assay exists. This approach would also allow for the identification of participants with prior immunity. To detect acute infection by PCR or other antigen-detection assay, this approach would require regular testing of all participants with or without clinical symptoms. Frequent specimen collection and testing would greatly increase the cost and logistical complexity of the trial. There may also be increased missingness in the data where burden on participants is high.

8.2.1.4 Disease Severity or Complication of Greatest Interest

In some settings, the public health outcome of interest may be in reducing severe disease or a complication of greatest interest. The primary endpoint could be restricted to a particular complication, such as laboratory-confirmed disease requiring hospitalization or congenital Zika syndrome in the children of women infected with Zika during pregnancy. The greatest expected challenge with these outcomes is that they may be very rare, resulting in inadequate power to test for a vaccine effect. The complication of greatest interest may also take a long time to develop, resulting in excessively long time trial scales. A replacement endpoint, such as an early complication associated with the endpoint of greatest interest, may be considered. This type of approach was recommended for HPV vaccine trials (Schiller et al. 2012).

Alternatively, one could assess the impact of the vaccine on reducing burden of illness as measured by a severity score, or a continuous measurement, such as viral load. In the case of viral load, there should be satisfactory evidence that it is well-correlated with disease severity for the pathogen of interest. Chang et al. (1994) suggested a general measure of efficacy that takes into account both the incidence of disease and severity. In practice, it may be challenging to objectively define a meaningful score. Furthermore, one needs to be sure that the analysis is not confounded by differences in the quality of medical care provided to infected trial participants (Hudgens et al. 2004). Over time, care may improve during the outbreak if cases are detected more quickly, understanding of the clinical phenotype advances, or the standard of care improves. The protocol must also clarify if the analysis will include in the denominator only those infected or everyone in the trial (Hudgens et al. 2004). The former approach would not be recommended as it is subject to selection bias because it is a post-randomization event and does not account for disease cases that were prevented by vaccine. The results of such an analysis could be misinterpreted as showing vaccine-enhanced pathogenesis when the vaccine actually prevents infection with more innocuous strains.

8.2.2 Estimating Intention-to-Treat and Per Protocol Effects

It is widely recognised that the primary analysis of a randomized trial should be an 'intention-to-treat' (ITT) analysis, in which all participants are analysed in the intervention group to which they were randomly assigned, regardless of the intervention received. In vaccine trials, an ITT analysis typically includes all cases occurring after randomization or all cases occurring after the first dose of vaccine/control (M E Halloran et al. 2010:6.3.1). ITT analyses are preferred because, providing that sufficiently many patients are randomized, the intervention and control groups will be balanced with respect to both measured and unmeasured baseline covariates. An ITT analysis estimates the effect of assignment to intervention or control, regardless of the proportion of trial participants who accept their randomized assignment or adhere to the intervention as specified in the trial protocol. An ITT analysis requires that outcomes are ascertained in all participants, and loss to follow up may lead to bias. White et al. (2011) discuss approaches to dealing with missing outcome data in ITT analyses, including the use of sensitivity analyses. Strategies to reduce and account for missingness are discussed in **Section 10.3**.

In contrast to other types of randomized trials, vaccine efficacy trials have typically adopted a per protocol approach that restricts the population to eligible, fully compliant participants receiving all doses as allocated per protocol (Horne, Lachenbruch, and Goldenthal 2001). The goal of the per protocol analysis is to estimate the per protocol effect, which is defined as the effect of intervention

(compared with control) that would be observed if all patients adhered to the trial protocol (Hernán and Robins 2017). Note that the trial protocol would mandate cessation of vaccination in patients who experienced an allergic reaction to the first dose of vaccine – even if such patients are incompletely vaccinated they have nonetheless adhered to the trial protocol. By contrast, a patient who declines initial vaccination because they have changed their mind about receiving an injection has not adhered to the trial protocol.

Unfortunately, standard per protocol analyses for estimating per protocol effects can be seriously biased. The potential for bias arises because adherence to study protocol, including receipt of all doses of vaccine, may be related to prognostic factors. Standard analyses may fail to capture losses to follow-up in key risk groups or due to important side effects of vaccination. This can be a major problem in settings in which a large proportion of the study population is noncompliant (Gilbert, Shepherd, and Hudgens 2013). See **Section 10.3** for general guidance on reducing loss to follow-up and missing data. Causal inference methods can be used to improve the quality of per protocol analyses. In particular, instrumental variable methods can estimate per protocol effects while avoiding the bias that arises because of confounding by prognostic factors that predict receipt of randomized intervention, including failure to receive all doses of a vaccine (Hernán and Robins 2017).

Another key feature of the per protocol analysis in vaccine trials is that events are only counted if they fall within an analysis period that typically begins after the last dose plus the maximum incubation period and some additional time for immunity to develop; early cases of disease or infection are not counted because they are considered attributable to infections occurring before vaccination or before immunity developed. The definition of analysis period should be documented in the trial analysis plan in advance of unblinded data being available and should have a clear rationale based on knowledge of the target disease. Examples of the per protocol analysis period in recent vaccine efficacy trials include 14 days after the last dose of a two dose bivalent killed whole-cell oral cholera vaccine (Bhattacharya et al. 2013), 28 days after the last dose of a three dose dengue vaccine (Capeding et al. 2014), and 10 days after randomization for a single dose Ebola vaccine (Henao-Restrepo et al. 2017). Considerations for the selection of this analysis period are described in (Dean, Halloran, and Longini 2016). Notably, a conservatively long waiting period after the last vaccine dose is best for reducing bias due to infections occurring before development of protective immunity, but this approach may negatively impact study power, especially in settings with low disease incidence.

Individuals who are randomized should meet the trial's eligibility criteria. However some randomized individuals may subsequently be found not to meet eligibility criteria, for example if a sensitive test for infection is available, but the test results are not known at the time of randomization. Post-randomization exclusion of individuals who test positive at baseline (Gilbert et al. 2011) should not introduce bias provided that exclusions are based on pre-randomization characteristics and are done in advance of unblinded data being available. This third approach, known as modified intention-to-treat (mITT), has been applied to HIV vaccine trials, and works well for diseases with long incubation periods.

When the vast majority of trial participants are fully compliant and have adequate follow-up, both ITT and PP approaches will provide similar conclusions (Horne et al. 2001). In general, the ITT estimate will tend to be diluted compared to the per protocol estimate for two reasons: (1)

deviations from randomized assignment and non-adherence will bias the estimated effect towards the null, and (2) the ITT estimate includes cases occurring in individuals who did not receive the full vaccination schedule, or cases occurring shortly after vaccination. The difference between the ITT and per protocol estimates of vaccine efficacy may be especially large in the setting of an outbreak if many infections occur during this per protocol delay (Dean et al. 2016), or if many of those randomized do not receive their assigned treatment. ITT analyses, unlike per protocol analyses, may capture early negative side effects of vaccine resulting in incomplete vaccination. By the same measure, ITT analyses may also capture post-exposure prophylactic effects of the vaccine if present. Examining outcomes stratified by time can help to disentangle these effects, providing that the sample size is adequately large.

ITT yields a practical estimate of vaccine effectiveness because it includes cases that may have been infected before the vaccine induced an immune response, as well as individuals who fail to comply with protocol, potentially for reasons relating to the vaccine itself; these variabilities may be context-specific, though, and thus hard to generalize to other settings. ITT estimates provide information about the effectiveness of an intervention/public health strategy using the vaccine, and because they reflect the speed at which the vaccine becomes protective, they may be more meaningfully compared across vaccines with different numbers of doses or ramp-up periods. In summary, it is recommended that both ITT and per protocol estimates of vaccine efficacy are reported (Horne et al. 2001): these reflect different target quantities (estimands) of interest (International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) 2014). In studies where compliance with protocol is not high, estimates of per protocol effects should use statistical methods that avoid, as far as possible, the potentially serious bias that can be associated with traditional per-protocol analyses that are restricted to individuals who receive the intervention as randomized.

8.2.3 Secondary Endpoints

Vaccine studies typically specify multiple pre-defined secondary endpoints. These may include endpoints as described in **Section 8.2.1** that were not selected as the primary endpoint, such as infection or a specific disease complication. Endpoints may be defined using different case definitions or different requirements for laboratory confirmation. For severe disease, death (cause-specific or all-cause) is a common secondary endpoint. As death is typically a rare outcome, studies will likely not be powered to detect a vaccine effect on mortality.

There may be interest in estimating vaccine efficacy to reduce infectiousness, though this quantity is only estimable from certain limited types of trial designs that include transmission networks, such as households. Examples of vaccines designed to reduce infectiousness include a pertussis vaccine in Senegal (Preziosi and Halloran 2003) and malaria transmission-blocking vaccines (Carter 2001).

Another secondary outcome of interest may be the durability of protection, where vaccine efficacy is assessed at different time points (Hudgens et al. 2004). The ability to assess this outcome will depend on the length of follow-up relative to the rate of waning protection.

8.2.3.1 IMMUNOLOGICAL ENDPOINT

Trials are strongly encouraged to collect immunogenicity data along with efficacy data in order to identify correlates and potential surrogates of protection (see **Section 7.2**). Such data would typically include a baseline blood sample and post-vaccination sample(s) to measure the immune response. Surrogates can only be reliably established from prior efficacy trials, but if good surrogates already exist, they might be useful as replacement endpoints to reduce the study size and timeline. Surrogates could also theoretically be identified in later trials and then applied to specimens stored during an earlier trial. If the outbreak ends while the study is ongoing, immunogenicity data may be the only available data to predict vaccine efficacy. Without an established surrogate of protection, though, immunogenicity data alone would not be sufficient to determine efficacy.

Baseline immunogenicity data is also important for assessing susceptibility of the study population and for examining the relationship between prior immunity and vaccine efficacy. The vaccine may appear less effective if many study participants have naturally induced immunity, so an available measure of baseline immunity is valuable for interpreting study data in settings where prior exposure to the pathogen is high. In addition, the relationship between prior immunity and vaccine efficacy can be complex, as was observed for a live attenuated dengue vaccine for which seronegative vaccine recipients experienced increased risk of severe dengue illness and hospitalization (WHO 2016b). This may be particularly relevant for studies of Zika and other flaviviruses. The analysis can be stratified on baseline seropositivity.

In practice, collecting biological samples may be challenging in epidemic settings. Study staff may be at increased risk of accidental needle sticks. Transporting, storing, and testing samples will increase cost and complexity of the trial. Operational and implementation considerations are discussed further in **Section 11**. For feasibility reasons, studies may consider planning an immunological sub-study with detailed follow-up of a subset of patients.

If the vaccine is found to be safe and efficacious, trials may choose to offer vaccination to all participants who are alive and uninfected at the end of the study and who did not receive vaccine during the study. Immunological samples collected at the time of this “close-out” vaccination may provide additional data that can be used to assess potential correlates of protection (Qin et al. 2008).

8.2.4 SAFETY ENDPOINTS

Assessment of vaccine safety and tolerability begins in Phase 1 and 2 trials and continues in Phase 3 trials. In emergency settings, earlier studies may have fewer safety data available to signal key safety endpoints to monitor in Phase 3 trials. The type of vaccine, including delivery mechanism, and its mechanism of action influence the methods and measurements used in safety assessment. Prior experience with similar vaccine formulations used for other diseases may inform safety endpoints selected for monitoring. Phase 3 trials may not be adequately powered to detect rare adverse events, in which case post-licensure surveillance will be important for identifying the range of adverse effects.

8.3 STUDY POPULATION

The population enrolled in the vaccine study impacts the study size, likelihood of success, and generalizability of the results. Considerations for selection of this population are described below.

8.3.1 General Population

Vaccine studies are typically implemented at sites with established high rates of disease transmission and draw participants from a general population representative of those who would ultimately receive the vaccine (the target population). Table **1** summarizes the preferred target population for a selection of the Blueprint priority diseases.

In contrast to endemic diseases, for emerging pathogens, areas with reliably high incidence may not exist. Areas with previously high incidence due to an outbreak may be poor candidates for study sites if few susceptible individuals remain. In fact, there is a risk that some or all selected sites will observe no transmission during the trial period, especially if transmission is highly localized and difficult to predict. Nonetheless, sites could be selected from areas hypothesized to be at adequately high transmission risk to support a vaccine study, and site selection could be informed by region-specific disease transmission models. Sites could be selected for reasons of convenience, such as high density population centres, areas with a history of running clinical trials, or areas with greater laboratory infrastructure.

8.3.2 High Risk Population

Vaccine studies may be targeted to prioritize enrolment of participants at high risk of infection, based on the specific epidemiology of the pathogen and the nature of the outbreak. Epidemiological studies from previous and current outbreaks would inform understanding of the types of populations at highest risk. Risk could be defined by occupation. Examples include health care workers (HCWs), abattoir workers, first responders, and front-line caregivers (i.e., doctors, nurses, janitors, people who collect bodies, and gravediggers). Risk could also be defined by behavioural factors, such as injection drug use or commercial sex work.

As the incidence of the primary endpoint in a high risk population is expected to be higher than in the general population, a smaller overall sample size would be required. Thus, a high risk population approach may be desirable if there are limited doses of vaccine available. The feasibility of this approach will also depend on the size and accessibility of the targeted population. While the sample size may be smaller, the time required to recruit individuals with a narrow set of characteristics may not necessarily be less than the time to recruit a larger sample with less restrictive entry criteria. If there are very few abattoir workers in the study area, for example, it may not be possible to meet the sample size requirements of a Phase 3 trial, or eligible participants may be geographically scattered in such a way that makes them logistically inefficient to target. One strategy to increase the sample size and decrease spatial scatter might be to define a trial population as a particular risk group plus their families or neighbours. If infection risk is difficult to predict by occupation or behavioural factors, a general population approach may be easier to justify and implement.

The approach of enrolling participants at highest risk of infection may also suffer from poor generalizability (poor external validity) if the study population differs from the target population that

would ultimately receive the vaccine (Hayes and Moulton 2008:38). For example, the vaccine may be targeted for all individuals 18-65 years of age, but few front-line caregivers may be over the age of 50 years.

8.3.3 Responsively Defined Population

For studies of emerging pathogens, there is a potential benefit in allowing the study population to be flexibly defined so that it may follow an outbreak as it develops. An outbreak moving into a new region may trigger the addition of a new study site in that area. In a ring vaccination trial, laboratory confirmation of a new case can trigger the definition of a new "ring" comprised of family members, contacts, and/or neighbours of the confirmed case (Ebola ça Suffit Ring Vaccination Trial Consortium 2015). See **Box 1** for more information on the ring vaccination trial design. Similar to the high risk population strategy, a responsively defined population may experience a higher incidence rate than the general population, resulting in a design requiring a smaller sample size.

This approach relies on a sensitive and fast surveillance system to inform the study in real-time, and the capability of rapidly establishing the apparatus for enrolling, testing and following individuals in new sites. Surveillance systems may need to be strengthened in potential study areas to maximize the probability that local cases are detected. While sites selected for general population studies may have been chosen for reasons of convenience, sites added responsively may be less optimal for a vaccine study (e.g., poorer laboratory infrastructure, staff strained by response to ongoing outbreak). Alternatively, the study could employ a team that travels to enrol new participants, rather than training local staff at each new site. The Ebola ca Suffit ring vaccination in trial used mobile teams to conduct participant enrolment, vaccination, and follow-up (Henao-Restrepo et al. 2015, 2017).

A responsive strategy works best for diseases that spread slowly through predictable contact networks and for fast-acting (ideally single dose) vaccines. Diseases that spread quickly may not afford the study staff adequate lead time to define the site, train local staff, recruit participants, and vaccinate. Individuals may already be infected at the time of vaccination. To increase lead time, study sites could be expanded to include lower risk individuals in the same contact networks who may contract the disease after multiple generations of transmission.

To further increase lead time, the population could be partially predefined, with preregistration of vaccination sites that are only fully vaccinated following the first signs of an outbreak. A small number of individuals could be vaccinated at these sites during preregistration to increase the likelihood that they are maximally protected by vaccine at the time of exposure. Persons vaccinated well before exposure to the pathogen will provide valuable immunological data that is different from data from persons vaccinated during the outbreak itself. Vaccine efficacy is also more straightforward to estimate and interpret in a pre-vaccinated population.

8.3.4 Special Populations

Certain human participants are categorized as vulnerable populations and require special treatment with respect to safeguards of their well-being (World Medical Association 2008).

The vulnerable individuals' freedom and capability to protect one-self from intended or inherent risks is variably abbreviated, from decreased freewill to inability to make informed choices. The vulnerable populations refers to but not limited to children, minors, pregnant women, fetuses, human in vitro fertilization, prisoners, employees, military persons and students in hierarchical organizations, terminally ill, comatose, physically and intellectually challenged individuals, institutionalized, elderly individuals, visual or hearing impaired, ethnic minorities, refugees, international research, economically and educationally disabled and healthy volunteers (Shivayogi 2013). All vulnerable groups and individuals should receive specifically considered protection.

Ethically acceptable research ensures that no group or class of persons bears more than its fair share of the burdens of participation in research. Similarly, no group should be deprived of its fair share of the benefits of research; these benefits include the direct benefits of participation (if any) as well as the new knowledge that the research is designed to yield.

Investigators, sponsors or ethical review committees should not exclude women of reproductive age from epidemiological research. The potential for becoming pregnant during a study should not, in itself, be used as a reason for precluding or limiting participation. However, a thorough discussion of risks to the pregnant woman and to her fetus is a prerequisite for the woman's ability to make a rational decision to enroll in an interventional study (Council for International Organizations of Medical Sciences (CIOMS) in collaboration with WHO 2009).

Exclusion of pregnant women is still usual practice in trials that do not address obstetric conditions, largely due to concern about birth defects after specific drug exposure in utero and the view that high fetal risk without important medical benefits for the mothers is not acceptable. Exclusion, therefore should not apply to women with life-threatening diseases, as illustrated by early HIV/AIDS drug trials which included pregnant women in the earliest phases – before completion of animal reproduction studies- because any risk to the fetus was balanced by an overwhelming potential benefit (prolonging life) to the mother. Results from studies that excluded pregnant women cannot be automatically extrapolated to pregnancy. This lack of data specific to pregnant women will negatively impact in the health of pregnant women and their access to interventions in future outbreaks. A disease with high fetal/neonatal death without intervention, investigational treatment of the mother could not place the fetus at "greater than minimal" added risk (Gomes et al. 2017).

When the social value of the research for pregnant or breastfeeding women or their fetus or infant is compelling, and the research cannot be conducted in non-pregnant or non-breastfeeding women, a research ethics committee may permit a minor increase above minimal risk. Short-term and long-term follow-up of the fetus and the child may be required in research involving pregnant and breastfeeding women depending upon the study intervention and its potential risks (Council for International Organizations of Medical Sciences (CIOMS) in collaboration with WHO 2016).

Participants aged less than 18 years (equivalent to age of majority) are considered minors and worldwide are not permitted to provide consent. There are two contradictory ethical requirements that need to be balanced with respect to children involved in research; (1) there is a need to obtain evidence of efficacy and safety for the medication(s) and (2) there is the need for respect and protection of the child in the research environment. To achieve this end, GCP in paediatric research demands that studies comply with the Declaration of Helsinki, ICH topic E11, European Union Directives and other relevant international guidelines. To justify any research project one must

balance the benefit/risk ratio, provide experienced, competent personnel and infrastructure, obtain adequate informed consent/assent, and have the study evaluated and approved by an ethics committee containing expertise on the rights and needs of children (Gill 2004).

Defined research entry criteria ensure clinical safety to vulnerable participants (Shivayogi 2013).

8.4 Comparator Arm

We strongly recommend the use of randomization in vaccine studies. A detailed discussion of the importance of randomization is provided in **Section 9.1.1.** In this section, we review options for the comparator arm in randomized trials. These include placebo or active control (**Section 8.4.1**), delayed vaccination (**Section 8.4.2**), and another vaccine candidate for the same target pathogen, as used in studies of non-inferiority (**Section 8.4.3**).

8.4.1 Placebo or Active Control

The most common choice for a control arm is a placebo or active control. An active control might be a different vaccine already licensed for the indication being studied, or it might be a licensed vaccine for some other indication that does not affect the probability of the study endpoint(s) and thus functions in the same way as a placebo for purposes of assessing efficacy. In this section, we exclusively consider the latter definition, where the active control is a vaccine for an unrelated but geographically relevant disease. The former definition, where the active control is a vaccine aimed at the same pathogen, is discussed in **Section 8.4.3**.

Trials with a placebo or active control arm will have the most straightforward analysis and likely the greatest power for assessing vaccine efficacy. The use of a placebo or active control is also typically coupled with blinding (or masking) which further strengthens the quality of results. Blinding can occur at many levels, including blinding of participants, blinding of clinicians, blinding of outcome assessors, and blinding of statisticians. Thus, even in studies that appear infeasible to conduct in a blinded manner, it may be possible to arrange blinded assessment of suspected illness or infection. Blinding is important for limiting bias, such as selection bias, detection bias, and performance bias. As it can be hard to replicate a vaccine that burns/stings like an active vaccine, the use of an active control instead of a placebo may make blinding more achievable.

A lack of blinding can lead to different behavioural changes across the two study arms that can induce bias in estimated efficacy. Plans for addressing potential bias in unblinded studies should be laid out in advance. In the STRIVE trial of an Ebola vaccine candidate in Sierra Leone, there was concern that front-line workers in the vaccinated arm might volunteer or be assigned to higher risk duty; to reduce bias, the trial staff used questionnaires to measure use of personal protective equipment or changes in duties (Widdowson et al. 2016).

8.4.2 Delayed Vaccination

In particular settings, a delayed vaccination comparator may be selected instead of a standard placebo or active control comparator. Individuals/clusters are allocated to either immediate or delayed vaccination, with a delay between the two that is shorter than the typical duration of a trial. Delayed vaccination involves one-way crossover of participants and is thus related to the stepped wedge design described in **Section 9.1.4**. Such a trial may be more difficult (albeit not impossible) to blind (Nason 2016). If unblinded, the trial is subject to the standard sources of bias in unblinded trials. A delayed vaccination comparator was used in the Guinea Ebola ring vaccination trial (Henao-Restrepo et al. 2015) and in the STRIVE Ebola vaccine trial in Sierra Leone (Widdowson et al. 2016).

The motivations for delayed vaccinations are to improve acceptability, to provide vaccine to individuals in greatest need, and to potentially avert more cases and promote epidemic control if the vaccine is efficacious. However, if the vaccine is ineffective or dangerous, more people may be exposed to the vaccine than would be in a trial with placebo/active control. If safety signals are recognized before the delay period has elapsed, then vaccine would not be given to the delayed group, and no more people would be exposed than in a design with a standard placebo/active control comparator.

The delayed vaccination approach generally leads to a decrease in power because each individual contributes less person-time at risk to the analysis, and that latent adverse effects may be harder to identify. Delayed vaccination can also lead to bias in estimated vaccine efficacy if the per protocol analysis period (see **Section 8.2.2**) includes time where the immediate vaccination arm is unprotected by vaccine or the delayed vaccination arm is protected by vaccine. To reduce bias and improve interpretability of results, the length of the delay should be relatively long compared to the disease incubation period and the time required for the immune response to develop among vaccinated persons. Further guidance on how to define this analysis period in order to reduce bias is available (Dean et al. 2016).

8.4.3 Another Vaccine Candidate (Studies of Non-Inferiority)

When evaluating an experimental vaccine in a setting where an existing vaccine has been established to provide clinically meaningful benefit, an ethical approach would be to conduct an active comparator trial that compares the experimental and existing vaccines. If the existing vaccine has moderate/poor efficacy, a standard superiority trial could be used to assess whether the experimental vaccine has superior efficacy relative to the existing vaccine. If the experimental vaccine has advantages over the existing vaccine other than efficacy, such as having a more favourable tolerability or safety profile, being more convenient to store, transport, or administer, or less costly, it might be sufficient for the experimental vaccine to have similar rather than superior efficacy relative to the active comparator vaccine. In PHE settings, we may also prefer to have multiple vaccines with established efficacy to reduce the potential for supply chain issues during future outbreaks.

It is not possible to establish that the experimental and existing vaccines have the same efficacy. Hence, the formal approach in such settings is to conduct a non-inferiority (NI) trial designed to determine whether we can rule out that the efficacy of the experimental vaccine is ‘unacceptably worse than’ that of the existing vaccine. This requires the formulation of an evidence based ‘non-

inferiority margin' that would be minimum threshold for an unacceptable loss of efficacy (Fleming 2008; Fleming et al. 2011).

For those designing an NI trial, it may be tempting to select a large NI margin as this will reduce the required sample size and increase the likelihood of drawing a conclusion of noninferiority. However, the risk of large NI margins is that an existing vaccine established to provide clinically meaningful protection could be replaced by a meaningfully less effective intervention, otherwise known as "non-inferiority creep."

To justify a non-trivial margin, the existing vaccine needs to have high efficacy that has been precisely estimated in a setting relevant to that in which the NI trial is being conducted. This last criterion is often referred to as the 'constancy assumption'. There are many factors that invalidate the constancy assumption, in particular factors that would cause the existing vaccine to be less effective in the setting of the NI trial than it was in the previous trials that established its efficacy, such as a change in the circulating pathogen strain. A proper formulation of a NI margin needs to account for the unreliability of the estimates of the effect of the existing vaccine in the setting of the NI trial (since it does not have a placebo control arm), including accounting for the inherent uncertainty about the validity of the constancy assumption, and needs to be chosen in a manner to preserve an important fraction of the effect of the existing vaccine (Fleming et al. 2011).

NI trials often, though not always, require large sample sizes. If it is reasonable to anticipate that the experimental vaccine could be modestly better than the existing vaccine, then a NI trial of moderate size could be well powered to rule out a relatively small margin. If the NI trial yields highly favourable results for the experimental vaccine that not only rules out the NI margin but also the hypothesis of equality, then that trial could justify a conclusion that the experimental vaccine has efficacy that is superior to that of the existing vaccine.

In NI trials, efficacy of the candidate vaccine is not directly observable; instead relative efficacy is estimated as the relative reduction in disease risk or incidence by one vaccine compared to another. In some studies, it may be possible to collect outcomes on individuals who do not consent to or are otherwise ineligible for vaccination. The outcomes of these individuals may form the basis for a non-randomized (observational) comparison to estimate vaccine efficacy. This approach is described in a randomized, double-blind trial in Senegal of two pertussis vaccines; the trial included a prospective, nested case-contact study and a cohort study to derive an estimate of absolute efficacy for each vaccine (Simondon et al. 1997).

An open-label NI trial was conducted in India to compare an existing subcutaneous injection vaccine with an experimental aerosolized vaccine for the prevention of measles. Eligible children aged 9 to 11.9 months were individually randomized to one dose of either the injection or aerosolized vaccine. The primary end points were seropositivity for measles antibodies and adverse events 91 days after vaccination. The pre-specified NI margin was 5 percentage points. 1560 children contributed to the per protocol analysis, and the relative efficacy of the aerosolized vaccine to the injection vaccine was -9.2 percentage points (95% CI, -12.2 to -6.3), which was inferior by the pre-specified NI margin (Low et al. 2015).

9 MAJOR STUDY DESIGNS TO EVALUATE AN EXPERIMENTAL VACCINE

In this section we review the major study designs for the evaluation of experimental vaccines. We consider their applicability to outbreaks and PHEs. The major study designs for evaluating an experimental vaccine are divided into two categories: randomized trials (**Section 9.1**) and observational studies (**Section 9.2**).

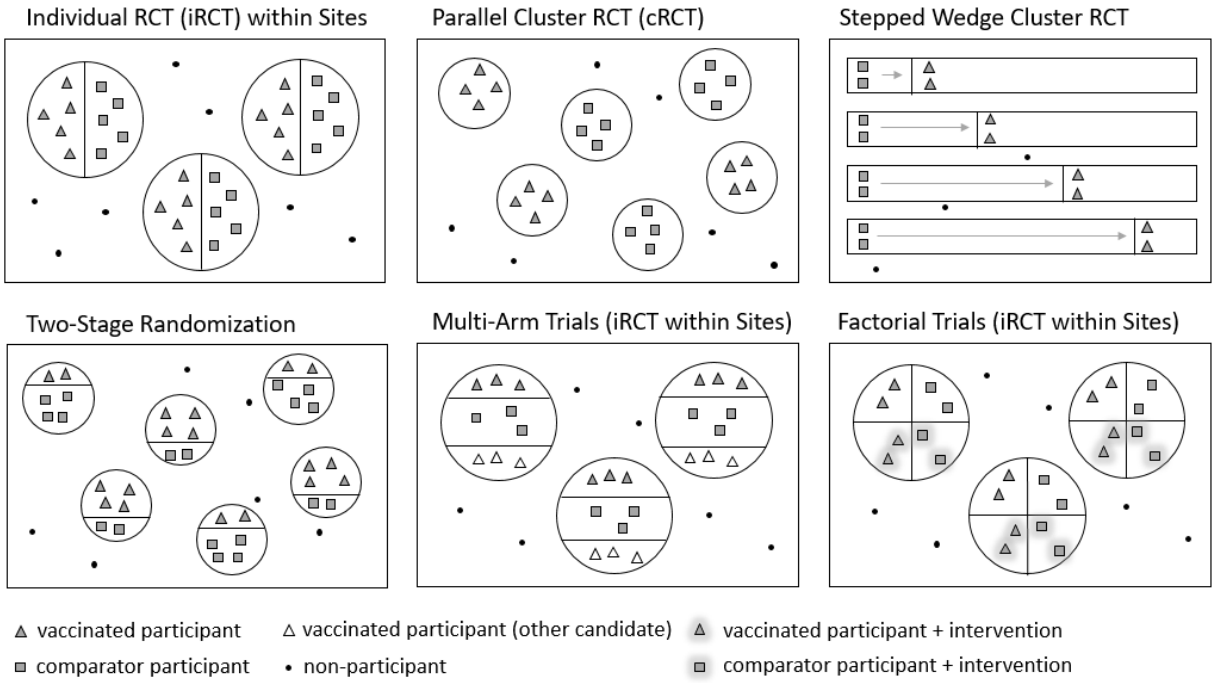
9.1 RANDOMIZED TRIAL DESIGNS

Randomized vaccine trials are evaluations in which the intervention is assigned via randomization. The objectives of such trials may be either to establish the efficacy of a single experimental vaccine against a control comparator (see **Sections 8.4.1 and 8.4.2**), to compare multiple vaccines (see **Section 9.1.5.2**), or to demonstrate that an experimental vaccine is comparably efficacious to a standard vaccine (i.e., a non-inferiority trial) (see **Section 8.4.3**).

We consider vaccine trials randomizing participants at the individual level (see **Section 9.1.2**), as a cluster of individuals (see **Sections 9.1.3** and **9.1.4**), or, less commonly, at both the individual and cluster levels ("two-stage randomization") (see **Section 9.1.5.1**). Cluster-randomized trials may employ a traditional parallel design (see **Section 9.1.3**), in which clusters are allocated to a treatment arm at the start of the trial and their allocation does not change, or a stepped wedge design, in which clusters starting in the comparator arm are vaccinated over time in a randomized order (see **Section 9.1.4**). These designs are summarized in **Figure 1.**

We also briefly describe several less commonly used randomized trial designs, including multi-arm trials (see **Section 9.1.5.2**), factorial trials (see **Section 9.1.5.3**), effectiveness trials (see **Section 9.1.5.4**), and human challenge trials (see **Section 9.1.5.5**).

FIGURE 1: SCHEMATIC OF RANDOMIZED TRIAL DESIGN OF RANDOMIZATION



9.1.1 ROLE OF RANDOMIZATION

The randomized clinical trial has traditionally been considered the gold standard study design for assessing efficacy of an intervention. The reason is that randomization provides assurance that the groups being compared are similar except for the assigned intervention. However, questions have been raised about the appropriateness of randomized designs during PHEs. These issues arise from concerns that randomization may deny persons an opportunity to have access to a potentially effective vaccine in a situation with high mortality and lack of adequate medical countermeasures. The design of vaccine trials in these settings poses complex questions, pitting issues of scientific rigor against feasibility and acceptability (Kanapathipillai et al. 2014). Questions have been raised about whether randomized designs would be acceptable to communities during outbreaks and the fairness of how a limited amount of product is being distributed.

The ethics of the use of randomization and placebos in vaccine trials as well as during public health emergencies have been considered by multiple organizations and advisory groups. These include the WHO Expert Consultation on the Use of Placebos in Vaccine Trials (World Health Organization 2013b) and WHO Guidance for Managing Ethical Issues in Infectious Disease Outbreaks (World Health Organization 2016). Much has been written about the Ebola experience in particular (Haire and Folayan 2016; Rid et al. 2014; Saxena and Gomes 2016; WHO Ethics Working Group 2014).

A criterion for placebo randomized controlled studies is equipoise. The equipoise criterion is normally met if trials will collect data needed to support licensure (generally, implying that there is still sufficient uncertainty about vaccine efficacy and safety), permitting use of placebo controls. If there is already sufficient efficacy and safety data to support licensure or if EUA/EUAL is used, then placebo-controlled trials may indeed raise ethical issues (see **Section 7.1**), but so would any trial that involves withholding an intervention known to be effective from some participants. In the limited settings where placebo-controlled trials are not considered acceptable, historical control or another

concurrent non-randomized control may be adopted if there is sufficient confidence that the rate of disease in the current outbreak is well estimated by historical rates, a requirement that is often difficult to meet. These are observational designs, which are considered further in **Section 9.2**.

These issues have been reviewed and weighed in in a recent consensus committee report of the National Academy of Medicine (NAM) (National Academies of Sciences Engineering and Medicine 2017). On balance the NAM Committee concluded that randomized controlled trials are the most reliable and rapid way to identify the relative benefits and risks of investigational products. The NAM Committee argued that every effort should be made to implement designs with random group assignment during outbreaks and epidemics and that only in rare circumstances should that not be possible. The WHO Working Group concurs with the conclusion of the National Academy of Medicine Consensus Report. Randomized trials are the study design of choice in PHEs and deviation from use of randomized designs should occur only under very exceptional circumstances. The role of observational studies is discussed in **Section 9.2.1**.

9.1.2 Individually Randomized Controlled Trial

Description

Individually randomized controlled trials (iRCTs) are studies in which each individual participant is randomized to either the experimental vaccine group or a comparator group. As vaccine efficacy trials can require large sample sizes, most iRCTs for vaccine efficacy are multi-site (multi-centre) trials with a few large, pre-defined sites and individual randomization within each site, i.e., a one-stage design with randomization stratified within sites (see **Figure 1**). A less common application of the multi-site iRCT design is to include a larger number of small study sites, such as households or villages.

The key advantage of the iRCT design is that the overall balance of measured and unmeasured confounders is typically much better for an individually-randomized trial than for cluster-randomized designs. As a result, iRCTs nearly always require fewer total participants than cluster-randomized trials in the same populations. To further enhance balance across trial arms, researchers may use stratified randomization procedures with key individual-level covariates that predict infection, such as age, gender, or known risk factors (Friedman, Furberg, and Demets 1998:5).

Another important advantage of iRCTs is that they are easier to blind than cluster-randomized trials (see **Section 9.1.3** for a discussion of the challenges of blinding cluster randomized trials) (Puffer, Torgerson, and Watson 2003). One possible exception may be the setting where individual randomization is conducted within small units, such as a household, and the vaccine has distinctive side effects (Nason 2016).

Analysis

The analysis is traditionally handled with a standard comparison of two independent groups using proportions, rates, or time to event methods. For multi-site trials, individuals within sites may have similar outcomes (as compared to individuals across sites), so the analysis should account for within-

site correlation. Options for adjustment include, but are not limited to, a site-level analysis comparing the two arms within each site, a regression model with site added as a covariate (fixed effect), a regression model with a shared random effect term for each site (if the number of sites is sufficiently large) (Donner and Klar 2004), a stratified analysis using a Mantel-Haenszel procedure, or a conditional regression model treating site as a nuisance variable. Adjusting for site improves precision because there may be significant variability in disease incidence across sites. For non-linear regression models (e.g. logistic regression and Poisson regression), failing to adjust for site can induce bias (Hauck, Anderson, and Marcus 1998). These procedures, which assume independence across sites, may miss dependencies between sites if sites are linked as part of a larger transmission network. Computer simulations may be useful for planning or analysing data from trials where such dependencies are expected (see **Section 10.4**). Though likely underpowered, simple analyses can be used to detect heterogeneity in the vaccine effect across sites or subgroups.

The primary analysis returns an estimate of vaccine efficacy measured as the direct effect of vaccination (see **Section 8.1**), which is typically used to evaluate a product for licensure (see **Section 6.1**). Both vaccinated and unvaccinated participants are assumed to receive similar indirect protection from vaccine in iRCTs. It is generally not possible to estimate indirect vaccine effectiveness in iRCTs, so an iRCT design would be inappropriate for transmission-blocking vaccines that do not also have direct effects. To estimate a vaccine's effect on reducing infectiousness in an iRCT, it would usually be necessary to use a proxy measure such as reduction in viral shedding (Hayes and Moulton 2008:31). However, if there is sufficient variation in the vaccination rate across sites, it may be possible to obtain some preliminary estimates of indirect effectiveness (Perez-Heydrich et al. 2014). In two-stage randomized designs (see **Section 9.1.5.1**), sites are explicitly randomized to a range of vaccination rates in order to support estimation of indirect effects.

SAMPLE SIZE CALCULATION

To conduct power and sample size calculations, it is necessary to assume predicted disease incidence in the comparator arm, vaccine efficacy, and expected rates of loss to follow-up. Type I error is typically set at 0.05. Power is typically set between 80 and 90%. For a two-arm trial, individuals are typically randomized in a 1:1 ratio. For trials with time to event analyses, unequal allocation (e.g., 2:1 vaccine to control) may result in comparable power as equal allocation, thereby reducing the number of participants required in the comparator arm (Sposto and Krailo 1987). Formulae for estimating sample size for iRCTs are available in Hayes and Bennett (1999) and Fay et al. (2007). Formulae are provided for comparisons based on population rates, proportions, or a continuous response endpoint. These are further summarized in Halloran et al. (2010:6.3.5).

For comparisons based on population rates, sample size calculations can be used to determine the required number of events or the required sample size to achieve the desired level of trial power. With the fixed number of events approach, calculations are based on type I error, power, and assumed vaccine efficacy. Participants are enrolled and followed until the required number of cases is reached. If the assumed disease incidence is incorrect, the study may take a longer or shorter period of time to complete, but power will not be affected. With the fixed sample size approach, the assumed disease incidence and loss to follow-up rates are used with the calculated number of events to determine the overall sample size required for the trial. Participants are enrolled until the

sample size is reached. With this approach, it is possible to end up with an insufficient number of endpoints at the end of the study if the assumption about disease incidence is incorrect.

The sample size calculation methods for iRCTs assume independence of participants, though this assumption may not be appropriate in the context of infectious disease trials. There may be transmission networks that create dependencies, or there may be indirect effects of vaccination that alter risk to other trial participants (see **Section 8.1**). As a result, standard methods could lead to underpowered trials. Simulation studies may be also be worthwhile for estimating power because these dependencies, as well as other complex factors, can be incorporated into the simulation (see **Section 10.4**).

EXAMPLES

The efficacy of a recombinant, live, attenuated, tetravalent candidate dengue vaccine (CYD-TDV) was evaluated in two large individually-randomized multi-centre Phase 3 trials. One trial was conducted at twelve centres in five Asian Pacific countries (Capeding et al. 2014), and the other trial was conducted at twenty-two centres in five Latin American countries (Villar et al. 2014). In both trials, healthy children were individually randomized in a 2:1 ratio to receive three doses of vaccine or placebo at 0, 6, and 12 months. Participants were followed using active surveillance for 25 months following the first dose. The primary analysis was estimated vaccine efficacy against symptomatic, virologically-confirmed dengue occurring between months 13 and 25 measured per protocol.

In another trial, the efficacy of intra-nasally administered live attenuated influenza vaccine was compared to the efficacy of inactivated vaccine in infants and young children (Belshe et al. 2007; Cox and Bridges 2007). The study was conducted at 249 sites (physicians' offices and primary care clinics) in 16 countries. On average, 34 children (range, 1 to 270) underwent randomization at each study site. Participants were randomized in a 1:1 ratio to live attenuated vaccine or inactivated vaccine, stratified according to age, previous influenza vaccination, history of recurrent wheezing, and country of residence. The primary analysis was estimated relative vaccine efficacy against culture-confirmed influenza-like illness, powered to detect superiority of the live attenuated vaccine.

CONSIDERATIONS IN OUTBREAKS AND PHES

One way that individually randomized trials can be adapted for outbreak settings is via the definition of the study sites. Typically vaccine trials enrol several large study sites defined as hospitals and their catchment areas; the advantage of this approach is that it uses existing clinical, surveillance, and laboratory infrastructure. Like the influenza trial described that included physician's offices (Belshe et al. 2007), we can consider smaller units as study sites. Sites could be defined as natural groupings of people at high risk of infection (see **Section 8.3.2**), such as frontline workers at a single location. The study can target a high risk population but choose to leave inclusion criteria broad so that the participants adequately reflect the target population for the vaccine (Kennedy et al. 2016). Contacts of confirmed cases can be encouraged to enrol. Stochastic disease modelling may be useful for identifying populations or geographic regions at highest risk of infection. Sites could be risk-prioritized such that those determined to be at highest risk are randomized and vaccinated first, in order to maximize the accumulation of high-risk person time (Bellan et al. 2015).

Sites could be defined responsively, as described in **Section 8.3.3**. If sites are case-ascertained and defined from the geographic or social network surrounding the case, we refer to this design as a *ring vaccination trial* (see **Box 1**); the ring design refers to how sites are defined, and it does not specify whether individual or cluster randomization is used to assign vaccination.

When the site is very small, such as a small community or even a household, individual randomization may not be acceptable to participants. Some may argue that individually randomized trials spread the vaccine more effectively, and that family members may prefer that at least half of the people in the cluster get the vaccine, which is better than potentially none. Nonetheless, others may find it unacceptable to vaccinate some but not others within a small unit such as a household (Hayes and Moulton 2008).

Indirect protection from participants in the vaccination arm could dramatically reduce or, in extreme cases, shut off transmission in the comparator arm (Longini et al. 1993). A lack of endpoints in the comparator arm may prohibit the estimation of vaccine efficacy via direct effects (see **Section 8.1**). This effect is strongest when the indirect effects of the vaccine are high for participants within a trial site and/or the basic reproductive number of the pathogen is close to one. This may also be most likely when more participants are allocated to the experimental vaccine (e.g., 2:1 allocation). Thus, there is a balance between achieving high vaccine coverage in the trial and statistical efficiency. Simulations are recommended to predict these effects during the trial planning process (Hitchings, Grais, and Lipsitch 2016) (see **Section 10.4**).

9.1.3 Parallel Cluster Randomized Controlled Trials

Description

In parallel cluster randomized controlled trials (cRCTs), clusters of individuals are randomized as a unit to the experimental vaccine group or some comparator (see **Figure 1**). Unlike the stepped wedge cluster randomized trial design (see **Section 9.1.4**), clusters in a parallel cRCT are randomized to receive one or another of the interventions before the start of the study, and the intervention does not change until the end of the study.

The optimal cluster definition will be highly context dependent. Each cluster may be defined, for example, as a single community, village, household, worksite, school, or medical centre/hospital. The cluster may be also chosen to match the planned vaccination delivery system, if vaccination delivery will occur through health care clinics, schools, or households (M E Halloran et al. 2010:13.3.2). A useful guide for defining clusters in a cRCT is available elsewhere (Hayes and Moulton 2017:4). All eligible members of a cluster may be invited to participate in a trial, or only a subset; ideally, the subset would be a representative sample of the larger cluster population, perhaps randomly drawn.

Cluster randomized designs are most appropriate when clusters are well-defined, stable, self-contained, and non-overlapping. Movement or transmission between clusters is referred to as contamination, which decreases estimates of indirect or total vaccine effectiveness by increasing the similarity of disease incidence across trial arms. Strategies to reduce contamination include selecting well-separated clusters and using buffer zones (Hayes and Moulton 2008:58–64).

As noted in **Section 9.1.2**, a cRCT nearly always requires a larger overall sample size than an iRCT in the same population. This is due to the correlation in outcomes between individuals within the same cluster, leading to diminishing returns as the average cluster size grows but the number of clusters stays the same. Thus, the number of clusters, rather than the overall sample size or size of the clusters, drives the efficiency of the trial (Pladevall et al. 2014). Many trials include too few clusters to justify the inferential methods used in the analysis. A trial with many smaller clusters will generally be stronger than a trial with few large clusters.

Parallel cRCTs are prone to greater baseline imbalance than iRCTs because fewer units are randomized (Hayes and Moulton 2008:40). Imbalances can occur across cluster- and individual-level variables. Even if some post-hoc adjustment is performed using measured confounders, trial results may still be subject to greater criticism. Trialists may thus consider stratified randomization or matching using cluster-level covariates to reduce the chance of severe imbalance, thereby improving the power and credibility of the study (Hayes et al. 2000). Cluster-level variables used for stratification or matching should be sufficiently correlated with the outcome to merit the added complexity and lost degrees of freedom in the analysis (Donner and Klar 2004; Hayes et al. 2000). Common choices include cluster size and geographic area (Donner and Klar 2004); predictors of transmission, such as presence/absence of other interventions, could be used; recent incidence data or model-based predictions of future incidence, if available, should also be considered for trials of infectious diseases (Hayes et al. 2000). These variables must be available prior to randomization and be uniformly defined from cluster to cluster (Pladevall et al. 2014). Matching also has other limitations (Donner and Klar 2004; Hayes et al. 2000), including risking being unable to match certain clusters if too many matching factors are employed. Furthermore, if a single cluster must be excluded from the trial analysis, it may be necessary to drop the entire matched pair (Hayes and Moulton 2008:70).

A key advantage of parallel cRCT designs is that they may be logistically easier to implement than iRCTs. For example, when all in a cluster receive the same treatment, it is easier to ensure that each participant receives the correct intervention (Hayes et al. 2000; Pladevall et al. 2014). This administrative convenience can reduce costs (Doussau and Grady 2016) and also limit contamination due to participants accidentally receiving the wrong treatment (Hayes et al. 2000; Pladevall et al. 2014). Cluster randomization may facilitate informed consent at the community-level in addition to participant-level informed consent. Furthermore, cluster randomized designs may be preferred in settings where participants are unwilling to consent to individual randomization within a cluster unit, such as within households (M E Halloran et al. 2010:13.3) or areas where decisions are made at the community-level (Hayes and Moulton 2008). Randomizing different members of a small cluster, such as a family or village, to different arms could result in participant dissatisfaction or tensions, especially when the trial is unblinded (Nason 2016).

Parallel cRCTs may be blinded if comparator clusters are offered a placebo or active control vaccine (see **Section 8.4.1**). Blinding is important for limiting bias in a randomized trial. It can be harder to maintain the blind in a cRCT because the unblinding of any one participant can lead to the unblinding of all participants in the cluster. It can also be harder to implement blinding in a cRCT because many trials may enrol participants after randomization for logistical reasons. If possible, unblinded trials should seek to recruit and consent eligible participants prior to randomization to prevent foreknowledge of allocation and reduce selection bias (Puffer et al. 2003). Ideally, clusters

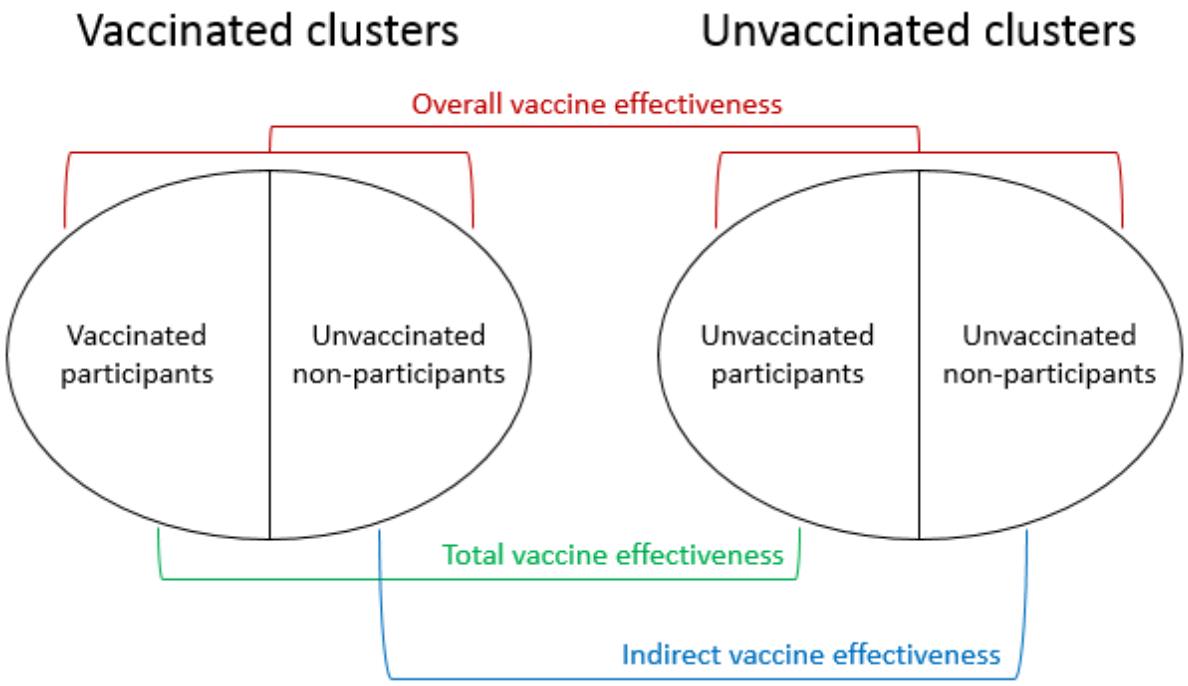
should have well-defined populations and be completely enumerated prior to study randomization and vaccination. To prevent performance bias (Higgins and Green 2011) and ascertainment/detection bias (Hróbjartsson et al. 2014), outcomes assessors and those providing medical care/treatment to trial participants should be blinded to allocation. There must be equal surveillance effort across clusters, especially if outcomes are hard to detect. Unlike iRCTs, some cRCTs with large clusters may compare population rates using primarily passive methods (e.g., reporting through existing surveillance) to detect outcomes; interpretability of the results will be limited by the quality of the surveillance system (Hudgens et al. 2004), and active surveillance methods are expected to yield more reliable data.

ANALYSIS

The analysis of a cRCT trial may be conducted at the cluster-level, by treating each cluster as the unit of analysis, or at the individual level, by treating each individual as the unit of analysis but properly adjusting for correlation between individuals within the same cluster. Cluster-level analyses can only adjust for cluster-level covariates, and there may be too few clusters to support this type of adjustment. Individual-level analyses can be adjusted for individual- and cluster-level covariates. Individual-level analyses must account for clustering of individuals outcomes. Options for adjustment include, but are not limited to, a mixed effects model with a cluster-level random effect (e.g., frailty) or a generalized estimating equations (GEE) model with a robust variance estimator. Adjusting for cluster decreases the precision of estimated vaccine efficacy, but analyses failing to adjust for cluster would be overly confident with inflated type I error (Pladevall et al. 2014). Detailed descriptions of the analytical methods for cRCTs is available in Hayes and Moulton (2017).

The primary analysis returns an estimate of vaccine efficacy measured as the total effect of vaccination (see **Section 8.1**) (see **Figure 2**). The total vaccine effect incorporates both direct protection from the vaccine and indirect protection from other vaccinated participants. Total vaccine effects are nearly always larger than direct vaccine effects, so the observed vaccine efficacy from a cRCT will be higher than from a comparable iRCT. It is possible to estimate indirect vaccine effectiveness by comparing incidence rates among unvaccinated individuals in vaccinated clusters (e.g., due to ineligibility or noncompliance) with incidence rates among individuals in unvaccinated clusters (see **Figure 2**). Estimating this quantity requires tracking the outcomes of individuals in clusters who are ineligible or otherwise do not consent to vaccination. This outcome is not estimated as rigorously as it would be from a trial that employed two stages of randomization – at the cluster level and individual level (see **Section 9.1.5.1**) – but this approach has gained some acceptance. Overall vaccine effectiveness, which compares the cluster-level incidence in vaccinated versus comparator clusters, can be measured in a similar manner (see **Figure 2**).

FIGURE 2: TYPES OF VACCINE EFFECTIVENESS MEASURED IN A CLUSTER RANDOMIZED CONTROLLED TRIAL. *Modified from framework developed by (Halloran et al. 1991). Vaccinated and unvaccinated refers to vaccination with the experimental vaccine. Unvaccinated participants may receive placebo, active control vaccine, or delayed vaccination.*



SAMPLE SIZE CALCULATION

To conduct power and sample size calculations, it is necessary to assume predicted disease incidence in the comparator arm, vaccine efficacy, and expected rates of loss to follow-up. Type I error is typically set at 0.05. Power is typically set between 80 and 90%.

For cRCTs, it is also necessary to assume the degree of clustering of outcomes in the population, as measured by the intracluster (or intraclass) correlation coefficient (ICC). The ICC measures the proportion of total variability in individual outcomes explained by the correlation between individuals in the same cluster. The ICC is used to calculate the trial design effect, which is a multiplicative factor that reflects how much larger a trial needs to be to achieve the same statistical efficiency as an iRCT. In its simplest form, the design effect for a cRCT is a function of the ICC and the number of participants per cluster (Hayes and Moulton 2008:7). For trials with complex designs, it may be more difficult to determine the design effect. A cRCT trial is least statistically efficient when the ICC is high and/or the number of participants per cluster is high. For a fixed number of clusters, the largest gains in power will be obtained when the number of participants per cluster size increases from 1 to about 1/ICC. Further increases will tend to have only a modest effect on power (Donner and Klar 2004). The cRCT design is often selected for its logistical advantages, and the cost (in terms of time or money) of enrolling a new site may be high relative to the cost of enrolling a new participant at an existing site. Cost-effectiveness equations to optimize cluster size given a fixed overall budget are available (Lohr 2010:193).

The ICC would need to be estimated from previously available studies or related data sets, where available (Ridout, Demétrio, and Firth 1999), but estimates can be very unstable, especially for trials with fewer than 40 clusters (Donner and Klar 2004). Since the ICC will not be known or will be estimated with uncertainty, calculations should examine sample size over a range of assumed values (Donner and Klar 2004). It is less widely appreciated that ICC is also likely to vary with cluster size (Hayes et al. 2000), with smaller clusters, such as households or small communities, expected to have higher ICC. Spatially resolved simulation models may be useful for assessing the relationship between ICC and cluster size (see **Section 10.4**).

Formulae for estimating sample size for cRCTs are available in Hayes and Moulton (2017). Formulae are provided for comparisons based on population rates, proportions, or a continuous response endpoint. These are further summarized in Halloran et al. (2010:13.6.1). Simulation studies may be also be worthwhile for designing parallel cRCTs (see **Section 10.4**). Examples of simulations used to design parallel cRCTs for HIV prevention interventions are described in Boren et al. (2014) and Wang et al. (2014).

EXAMPLE

The efficacy of a single dose of the Vi polysaccharide typhoid vaccine was evaluated in a Phase 4 parallel cRCT in slum-dwelling residents of Kolkata, India (Sur et al. 2009). The study area encompassing most of two wards in Eastern Kolkata was partitioned into 80 contiguous geographic clusters. Clusters were divided into eight strata according to ward, the number of residents who were 18 years of age or younger (<200 vs. ≥200), and the number of residents who were older than 18 years (<500 vs. ≥500). Using stratified randomization, 40 clusters were randomized to receive the Vi typhoid vaccine and 40 clusters were randomized to receive the comparator hepatitis A vaccine. Cluster members 2 years of age or older were targeted for vaccination, and vaccine coverage in clusters was about 60%. For the analysis of vaccine protection, Cox proportional hazards models were fit to the individual data, and standard errors were adjusted using a robust variance estimator. A set of analyses adjusting for the stratifying variables and other key individual-level covariates were also conducted. Total vaccine effectiveness was estimated to be 80% (95% CI, 53-91). Using outcomes from unvaccinated cluster residents, indirect and overall effectiveness were estimated to be 44% (95% CI, 2-69) and 57% (95% CI, 37-71), respectively.

CONSIDERATIONS IN OUTBREAKS AND PHES

For outbreaks and PHEs, a natural modification of the parallel cRCT design is to define clusters as groupings of individuals at high risk of infection (see **Section 8.3.2**); examples might include groups of individuals with a shared occupational exposure, such as workers in the same mine or frontline healthcare workers at the same clinic. The clusters ideally will capture transmission networks for the pathogen of interest (Hayes et al. 2000). Certain routes of transmission lead to well-defined clusters for which transmission primarily arises from infections within clusters versus from outside (e.g., sexual transmission, direct transmission, and transmission within households and schools). Nosocomial transmission may be paired with medical centres/hospitals as clusters. Vector-mediated transmission or respiratory transmission may be paired with a geographic cluster definition. Another

modification is to define clusters responsively, such as villages with at least one laboratory-confirmed case of the target disease (see **Section 8.3.3**). For more information about case-ascertained clusters, see **Box 1** on ring vaccination trials. To move some of the administrative burden before case-ascertainment, clusters could be pre-registered and receive enhanced surveillance. It may be difficult to match clusters in real time with responsive designs, but clusters could be block randomized with respect to a few key cluster-level covariates. This would help achieve balance in covariates as well as balance with respect to time of cluster enrolment. This strategy was used in the Ebola ça Suffit ring vaccination trial with cluster location (urban vs. rural) and cluster size as stratifying covariates (Henao-Restrepo et al. 2015).

One important disadvantage of parallel cRCTs in outbreak settings is that there may be surges of infections in different areas, and there may be heterogeneity in the uptake of the other interventions to prevent transmission. It is recommended that the trial track the use of alternative control measures in clusters as they provide external baseline information to improve estimation. These factors heighten differences in disease incidence between study sites and increase ICC, further increasing the sample size for cRCTs. If the cRCT is unblinded, additional resources might be deployed to comparator clusters not receiving the experimental vaccine, resulting in performance bias. The randomized comparison is thus a comparison of two intervention strategies (experimental vaccine versus enhanced infection control), but it is no longer a test of the isolated effect of the vaccine. Tracking the use of alternative control measures can help detect this type of bias.

In some scenarios, it may be necessary to exclude entire clusters in which data are expected to become unreliable. This could occur because of failure to deliver vaccine, failure of the cold chain, or other overwhelming logistical challenges (Lipsitch et al. 2015). If these types of losses to follow-up are expected, it is preferable to enrol many small clusters rather than few large clusters, to limit the overall impact on the sample size and precision.

BOX 1: RING VACCINATION TRIAL

Ring vaccination trials are modelled after the surveillance and containment strategy of ring vaccination, which was used in the 1970s to eradicate smallpox (Foege, Millar, and Henderson 1975). The concept for the ring trial was first detailed in (Ebola ça Suffit Ring Vaccination Trial Consortium 2015) to describe the design used for a Phase 3 Ebola vaccine trial in Guinea (Henao-Restrepo et al. 2015, 2017).

The key feature of the ring vaccination trial is how sites/clusters are identified and defined. Each site/cluster is linked to a confirmed disease case, identified through active or passive surveillance. The site/cluster is then defined as a "ring" around the confirmed case, representing people at risk of exposure to the pathogen. Rings can be geographically defined, contact-based (contacts and contacts of contacts), or some combination. The ideal choice of ring will depend on the networks by which the pathogen is transmitted (sexual, person-to-person, nosocomial, etc.). Rings should be clearly separated to limit contamination (Hayes and Moulton 2008:46). Individuals within rings can then be randomized as a cluster unit to an intervention in a cRCT (see **Section 9.1.3**), as originally described, or they can be individually randomized within the ring in an iRCT (see **Section 9.1.2**). The choice between the two strategies will depend on many factors, including the level of intracluster

correlation in the population, the level of indirect protection of the vaccine, the importance of estimating indirect and overall vaccine effects, the logistical feasibility of conducting individual randomization within rings, and the acceptability of conducting individual randomization within rings.

The key advantage of this trial design is that rings are responsively defined and follow the outbreak as it progresses. Ring trials would be expected to enrol more people at high risk of exposure than a trial with a predefined population if the outbreak is spatiotemporally unpredictable. These trials work best when the outbreak is highly localized, infected individuals and exposed contacts can be rapidly identified, and the vaccine works quickly enough to protect these exposed contacts. Another key advantage is that the design has naturally phased rollout, which can be valuable if resources are limited. This approach also shortens the lag between starting a trial and accruing sufficient high-risk person-time (Lipsitch et al. 2015). The key disadvantage of this approach is that it will not work well if the vaccine requires multiple doses or is otherwise slow to protect.

9.1.4 Stepped Wedge Cluster Randomized Trial

Description

The stepped wedge trial is a type of cluster randomized controlled trial (cRCT). Unlike parallel cRCTs (see **Section 9.1.3**), in which clusters are allocated to one treatment arm and then remain in that arm throughout the entire trial, in stepped wedge cRCTs, all clusters commence the trial in the control arm. The intervention is then introduced gradually at regular intervals during the trial, either one cluster at a time or in small groups of clusters, until by the end of the trial it is in place in all clusters (Hayes and Moulton 2008:136–41) (see **Figure 1**). The order of roll-out is randomized to address confounding and reduce bias. Stepped wedge trials are sometimes referred to as one-way crossover trials or phased implementation designs. General summaries of stepped wedge trials are available in (Hayes and Moulton 2008:136–41), (Copas et al. 2015), and (Hussey and Hughes 2007); a detailed summary on stepped wedge trials for infectious diseases is available in (Doussau and Grady 2016); a summary considering vaccination trials specifically is available in (Halloran et al. 2010:13.4.3)

The main reason for adopting this design is when there is already considerable evidence that the intervention may have a beneficial effect. As the intervention is presumed to do more good than harm, all participants receive the intervention at some point during the study (Doussau and Grady 2016). If the vaccine cannot be readily delivered simultaneously throughout a large area, either for logistical reasons or because of insufficient supply, random selection is a fair way to determine the order of rollout and allows for unbiased estimation of the treatment effect (Hayes and Moulton 2008:136–41). Stepped wedge designs can also be used when a parallel design with a placebo or active control comparator is infeasible either for practical or for ethical reasons, such as the vaccine already being licensed (M E Halloran et al. 2010:13.4.3).

Since stepped wedge trials are a type of cRCT, they retain some of the logistical advantages associated with implementing the intervention at the cluster-level (see **Section 9.1.3**). On the other hand, they also retain many of the disadvantages of parallel cRCTs. They require a larger sample size

than comparable iRCTs because of intracluster correlation. Stepped wedge cRCTs may require a larger sample size than even parallel cRCTs because of the imbalance they produce in the analysis between vaccinated and non-yet-vaccinated arms (e.g., at the beginning of the trial only one cluster is vaccinated and at the end of the trial only one cluster is unvaccinated).

Stepped wedge cRCTs are typically not blinded, and generally neither a placebo nor active control is used prior to vaccination with the candidate vaccine (M E Halloran et al. 2010:13.3.5.1). As a result, these trials may be prone to the types of bias associated with unblinded studies, such as selection bias, performance bias, and detection bias (see **Section 9.1.3**).

ANALYSIS

The analysis of stepped wedge cRCTs is more complex than the analysis of parallel cRCTs because clusters change treatment arm over time, precluding a standard two-arm comparison. A simple before versus after approach cannot be adopted either because of secular time trends. Disease incidence may be naturally declining over time or due to the presence of other interventions, especially in the context of PHEs, which can upwardly bias the estimate of vaccine efficacy.

The analysis typically takes either a horizontal or vertical approach to adjust for secular time trends. In the horizontal approach, time trends are explicitly modelled, potentially by examining changes in time for clusters where treatment status has not changed. This approach may require many degrees of freedom and may be more susceptible to model misspecification if, for example, no smooth trend exists (Hayes and Moulton 2008:253–54; Moulton et al. 2007) or if time trends are highly variable across clusters (Bellan et al. 2015). In the vertical approach, comparisons are only made within time steps and time is conditioned out as a nuisance parameter, accounting for within-cluster correlation (Hayes and Moulton 2008:253–54); one disadvantage of this approach is that it does not use all available data because the periods where everyone is unexposed and everyone is exposed are dropped from the analysis.

As participants within the same cluster are vaccinated at the same time, the natural estimate of vaccine efficacy from horizontal or vertical analyses is the total effect of vaccination (see **Section 8.1**). Indirect and overall effects can also be estimated from stepped wedge cRCTs if data is collected from cluster members who do not receive vaccination. The direct effect of vaccination cannot be estimated by randomized comparison in a stepped wedge cRCT.

SAMPLE SIZE CALCULATION

The design of stepped wedge trials can be complex. It is necessary to specify the following key design elements prior to implementing the trial: the size of clusters, the number of clusters receiving the intervention per step (as multiple clusters may crossover at the same time), the number of steps, the length of time between successive crossover points (step length), and the rollout period (baseline data collection before first crossover) (Copas et al. 2015). These design features collectively determine the overall sample size and trial duration.

To conduct power and sample size calculations, it is necessary to assume predicted disease incidence in the comparator arm, vaccine efficacy, the ICC, and expected rates of loss to follow-up. Type I error is typically set at 0.05. Power is typically set between 80 and 90%. As for parallel cRCTs, the sample size of a stepped wedge cRCT will be inflated by the trial design effect, which is a multiplicative factor that reflects how much larger a trial needs to be to achieve the same power as a comparable iRCT. Formulae for estimating sample size for stepped wedge cRCTs are available in Moulton et al. (2007) and Hussey and Hughes (2007). These are further summarized in Halloran et al. (2010:13.6.3). Simulation studies may be also be worthwhile for estimating power because even complex design features can be incorporated into the simulation and the assumption of constant incidence can be relaxed (see **Section 10.4**).

EXAMPLES

The stepped wedge cRCT design was successfully used in the Gambia Hepatitis Intervention Study (The Gambia Hepatitis Study Group 1987). This study evaluated the long-term effects of infant hepatitis B vaccination on preventing chronic liver disease and liver cancer. Seventeen vaccination teams were each assigned a portion of 104 delivery points that were visited at least once every two weeks to conduct routine immunizations. Every 10-12 weeks, a new vaccination team was instructed to introduce hepatitis B vaccine, with teams selected in a randomized order. After a four-year period, all delivery points included hepatitis B vaccine in routine vaccination. The statistical analysis used a vertical approach, dividing time into three month time periods and comparing outcomes for vaccinated and unvaccinated children. As the primary endpoints were long-term endpoints, over 20 years of follow-up have been conducted so far, and the trial is ongoing.

CONSIDERATIONS IN OUTBREAKS AND PHES

The best application of stepped wedge designs is for interventions that are already going to be implemented. Such an intervention would typically have previously established efficacy and safety. Randomization supports unbiased estimation of the intervention effect, and phased roll-out helps to prevent overburdening of resources. For an experimental vaccine, the use of a stepped wedge design for licensure would be unprecedented (Doussau and Grady 2016). There would need to be clear rationale/evidence to believe that the experimental vaccine would do more good than harm but often acquiring this evidence is exactly the goal of a Phase 3 trial. (See **Section 9.1.1** for an expanded discussion on randomization in PHEs.)

Though some have advocated for the use of stepped wedge cRCTs in outbreaks and PHEs (Piszczek and Partlow 2015), the design has important disadvantages, primarily related to the complexity in planning, implementation, and analysis (Hayes and Moulton 2008:254). All participants and facilities must be enrolled before the first dose of vaccine can be administered (Widdowson et al. 2016). Thus, there is no option to responsively add sites as the outbreak develops (see **Section 8.3.3**). These factors may delay the start of the trial, which can reduce overall power, especially in settings where the epidemic is declining (see **Section 10.1.1**). There are also limited or no options to expand the sample size, which could pose a problem if the disease incidence is lower than predicted (Widdowson et al. 2016).

Stepped wedge cRCTs are probably also the slowest trials, in part because of their complexity (Cohen and Kupferschmidt 2014). Step lengths should be long enough that the vaccine can induce protection within the step (Hussey and Hughes 2007); thus, trials with multi-dose or slow-acting vaccines may need to be very long, or the results may end up being difficult to interpret. Furthermore, we are unaware of any methods to implement interim analysis in a stepped wedge design (see **Section 10.1**), so there is no flexibility to stop a trial early for efficacy (Doussau and Grady 2016).

Finally, stepped wedge cRCTs are not well-suited for endpoints with spatiotemporally variable incidence (Bellan et al. 2015; Cohen and Kupferschmidt 2014; Doussau and Grady 2016). Infection risks may differ substantially across clusters and over time. There may be significant secular trends, such as waning of the epidemic due to its natural course or differential implementation of outbreak control strategies (Doussau and Grady 2016). It may be difficult to properly model these time trends in a horizontal analysis, and model misspecification can result in increased type I error (Bellan et al. 2015). There may also be seasonality in transmission, such as for vector-borne diseases. Stepped wedge cRCTs should avoid steps occurring during periods of no active transmission, which may make these trials challenging to operationalize for diseases with a low season.

In an effort to adapt this trial design to PHEs, several papers have presented modifications to the stepped wedge cRCT allowing clusters at highest transmission risk to be preferentially randomized to vaccination earlier in the trial (Diakite et al. 2016; Harling et al. 2017). Transmission risk may be predicted using a spatially structured mathematical and computational model (Diakite et al. 2016) or predicted using a network analysis (Harling et al. 2017). The goal is to deploy vaccine to participants at greatest risk of infection and to increase study power. These “stepped wedge family” designs are still expected to face many of the logistical and analytical challenges described above for stepped wedge cRCTs.

9.1.5 Other Trial Designs

9.1.5.1 Two-Stage Designs: Individual Randomization within Cluster Randomized Trials

In a two-stage randomized design, clusters are first randomized to some fixed level of vaccine coverage (e.g., low (20%) vaccine coverage or high (80%) vaccine coverage), and then individuals are randomized within each cluster to receive vaccine or comparator based on the coverage level determined in the first stage (see **Figure 1**). It would be possible to also randomize some clusters to no vaccine coverage or complete vaccine coverage, although this changes the nature of the design because individual randomization would not occur in those clusters.

This design is sometimes referred to as two-step or split-plot randomization (Hudgens et al. 2004), pseudo-randomization (Borm et al. 2005), or randomized saturation (Baird et al. 2016). “Two-stage” refers to the two stages of randomization (cluster and individual). It should be distinguished from trials that use an interim analysis to make a planned modification to the trial, such as dropping trial arms that are not promising, as these are also referred to as two-stage designs (Simon 1989). Seamless Phase 2/3 designs fit this latter category, and they are further discussed in **Section 9.1.5.2**.

An advantage of this design is that individual randomization reduces selection bias that occurs when participants are recruited and enrolled into a cluster with foreknowledge of the cluster's allocation. Furthermore, it is one of the only designs to support estimation of both direct and indirect vaccine effects (see **Section 8.1**) (Hudgens and Halloran 2008). For detecting major vaccine effects, two-stage designs are less powerful than iRCTs and parallel cRCTs; these standard designs offer a sharper contrast between trial arms, and power to detect indirect effects (e.g., through the use of a gradient of vaccination coverage) is detrimental to power to detect major vaccine effects (Baird et al. 2016). The other major disadvantage of this design is its complexity. As far as we know, no vaccine trials have employed this design. More information on two-stage randomization designs is available in Hudgens and Halloran (2008), VanderWeele and Tchetgen Tchetgen (2011), and Baird et al. (2016).

9.1.5.2 Trials with Multiple Vaccine Candidates and a Control Comparator Arm

Trials may be designed to include multiple vaccine candidates and a control comparator arm (placebo or active control vaccine) (see **Figure 1**). This design has an important advantage that the same trial infrastructure is used to evaluate multiple vaccine candidates and so may require fewer resources overall than multiple, independent two-arm trials (Parmar, Carpenter, and Sydes 2014). The design also facilitates direct comparison between the candidates because vaccine efficacy is measured concurrently in the same population using the same endpoints and methodology. This approach works best when the vaccines have complimentary mechanisms of action and similar target populations. Trials with multiple experimental vaccine arms are most commonly iRCTs (see **Section 9.1.2**) or parallel cRCTs (see **Section 9.1.3**) (Hayes and Moulton 2008:130).

The control comparator arm is typically shared by all experimental vaccine candidates, and vaccine efficacy is measured compared to the control comparator arm. Sharing the control comparator arm induces correlation between the measured vaccine efficacy of each experimental candidate. In the analysis, one must define if it is necessary to adjust for multiple comparisons, or if each candidate can be treated as a separate test with type I error of 0.05. This is especially challenging if the products are developed by different companies with different stakeholders.

Trials could include adaptive strategies to drop poorly performing candidates, though this approach is more typical of Phase 2 trials than Phase 3 trials. When determining the data monitoring strategy (see **Section 10.1**), the risk of excluding an effective vaccine candidate too early must be weighed against the risk of diluting power by splitting the population into many groups. Data monitoring strategies should further consider how early evidence of efficacy of one candidate impacts the continuation of the other active arm(s) and the control comparator arm. Early discontinuation of other arms would be risky if there are safety or supply issues with the first vaccine (Nason 2016). It may also not be clear which vaccines will ultimately be best-suited for different populations or circumstances (Osterholm et al. 2015) In this case, the data safety monitoring board (DSMB) may want to consider the conditional power estimates for the vaccine that does not have clear and substantial efficacy (Kennedy et al. 2016).

Phase 2 and Phase 3 trials may be formally combined into Phase 2/3 trials, also known as "seamless Phase 2/3", "discovery into confirmatory", or "combined-phased" trials. The first part of the trial is a smaller Phase 2 trial that collects data on safety and immunogenicity in a typically limited and focused study population, and the trial may use an "intermediate" efficacy endpoint. Some

preliminary assessment of efficacy might be provided. The second part of the trial is the larger Phase 3 trial that collects data on safety and vaccine efficacy. Analysis of the first phase provides a GO/NO GO decision for how/if to proceed to the next phase following a decision-making strategy clearly defined in the protocol. Trials may be inferentially seamless, in which data from the Phase 2 portion contribute to the Phase 3 analysis, or operationally seamless, in which the data are analysed separately. The advantages of this approach include operational efficiencies, with time potentially saved in writing protocols, finding sites, obtaining ethical and regulatory approvals, and ramping up enrolment of participants (Saxena and Gomes 2016). If the trial is inferentially seamless, an overall smaller sample size may be required than for two independent trials, though typically only minimal gains in statistical efficiency are expected. The disadvantages of this approach include that the trial may be more difficult to plan, contributing to delays in initiation of the Phase 2 component. There may be limited or no opportunity to make changes to the Phase 3 protocol, which is especially difficult in the setting of an emerging pathogen where diagnostics and understanding of the disease/vaccine may evolve rapidly. The results may be more difficult to interpret. It can be difficult to predict the financial and operational costs of the trial.

One implementation of the seamless Phase 2/3 design involves multiple vaccine candidates in Phase 2, with only the most promising being advanced to Phase 3. This type of adaptive clinical trial design has been extensively considered for non-infectious disease trials (Barker et al. 2009). It may be challenging to implement in the setting of emerging pathogens since not all vaccines may be available to start Phase 2 testing at the same time. They may also have differences in their profiles (number of doses, speed of action, eligibility criteria) that make it difficult to evaluate them together in a single trial.

The PREVAIL I trial in Liberia evaluated two active Ebola vaccine candidates (Kennedy et al. 2016). It was designed as an individually randomized, double-blind, placebo-controlled Phase 2/3 trial starting at a single site, focused on safety and immunogenicity, and with plans to expand to additional sites with a simpler efficacy endpoint in the Phase 3 portion of the trial. Two placebos were designed to match the volume of each candidate, and individuals were randomized 2:1:2:1 (vaccine A, placebo A, vaccine B, placebo B). The primary endpoint was definite Ebola infection occurring 21 days or more following randomization with separate analysis of each vaccine versus the pooled placebo group. As the trial approached its Phase 2 recruitment target, the outbreak in Liberia had waned such that it was deemed infeasible to collect an adequate number of primary endpoints to determine efficacy. Data on safety and immunogenicity were collected.

Gilbert et al. (Gilbert et al. 2011) described a Phase 2b trial design strategy for simultaneously evaluating multiple prime-boost HIV vaccine regimens against a shared placebo group in the same geographic region. Features of this design include the measurement of durability of vaccine efficacy for each regimen, the evaluation of immune correlates of protection, and the comparison of vaccine efficacy across regimens. The design uses sequential monitoring to drop vaccines with evidence of poor safety or efficacy. This trial design has not yet been implemented in the field.

9.1.5.3 FACTORIAL TRIALS

Factorial trials allow investigators to evaluate more than one intervention in a single experiment. For vaccine trials, factorial trials could be used to simultaneously evaluate an experimental vaccine

candidate and some other non-vaccine disease-prevention intervention, such as vector control for vector-borne pathogens or a strategy for behavioural risk reduction. The vaccine and other experimental intervention should have overlapping eligibility criteria, complementary mechanisms of action, and compatible toxicity profiles. Typically all possible combinations of each intervention are included as distinct arms in the trial (see **Figure 1**). More information on individually randomized factorial trials in Friedman et al. (1998:53-4) and on cluster randomized factorial trials in Hayes and Moulton (2008:8.2.2).

Factorial trials may be more efficient than two separate trials of the vaccine and the other intervention because they utilize the same population and trial infrastructure. In the absence of effect modification, a well-designed factorial trial will have the same power as two independent trials. Factorial trials can be used to assess effect modification of one intervention on the other, though generally power to detect these types of interactions will be low unless the sample size is dramatically increased to do so.

Disadvantages of factorial trials include that exclusion criteria must be appropriate for all interventions; the exclusion criteria used for a factorial trial thus may be need to be more restrictive than the exclusion criteria used for a trial only evaluating the experimental vaccine. Interventions may interact such that one intervention affects the compliance with others, making the results more difficult to interpret. Finally, factorial trials are more complex to design than standard vaccine trials.

A factorial trial was used to simultaneously evaluate the impact of vitamin A supplementation and vaccination against tuberculosis with the Bacillus Calmette-Guérin (BCG) vaccine to improve one-year mortality outcomes in low birthweight infants in Guinea-Bissau (Benn et al. 2010). Low birthweight infants at the participating national hospital were individually randomized to one of four arms, with either early BCG vaccination or standard, delayed BCG vaccination (unblinded), and with either 25,000 international units (IU) vitamin A or placebo (blinded). No interaction was observed between vitamin A supplementation and BCG vaccine allocation (p=0.73). The trialists were a priori primarily interested in the impact of the vitamin A supplementation program, and they noted no improvement in mortality for infants receiving vitamin A supplementation, and, worryingly, evidence that vitamin A supplementation at birth may be harmful in girls.

9.1.5.4 Effectiveness Trials

Vaccine efficacy can be distinguished from vaccine effectiveness, where the former is an estimate of the intrinsic vaccine effect measured in an idealized setting, and the latter is measured in a “real world” setting (Clemens et al. 1996). Vaccine effectiveness trials are population-specific trials that focus on estimating the public health impact of the vaccine under non-idealized conditions. For example, a country may not be able to adequately maintain a cold-chain, thereby reducing the effectiveness of the vaccine or it may not always be feasible to reach some parts of target populations. Results are not generalizable but could support country-specific licensure and provide useful information to local policy makers.

9.1.5.5 Human Challenge Trials

In human challenge trials, participants are intentionally exposed to the pathogen whether or not they have been vaccinated. The challenge pathogen may be close to wild-type and pathogenic, or it may be attenuated or genetically modified. These studies require extensive ethical review because of the risks posed to participants, and this study design would not be considered ethical or safe for many pathogens, especially those lacking efficacious treatments.

Human challenge studies can use classical experimental designs and relatively small sample sizes to directly assess efficacy, safety, and immunogenicity of an experimental vaccine. The close follow-up of participants also supports observation of many other vaccine effects that cannot typically be measured in vaccine efficacy trials (Hudgens et al. 2004). For example, these trials may be used to enhance understanding of pathogen and vaccine kinetics and to identify potential immune correlates of protection (see **Section 7.2**).

In rare settings, human challenge studies may be used to support regulatory decisions, provided that that the human challenge model system is adequately predictive of efficacy in the field. For example, in 1998 the FDA determined that human challenge would have been considered acceptable to demonstrate efficacy of a cholera vaccine (Food and Drug Administration 2011), and in 2015 licensed a cholera vaccine with data from Phase 3 challenge studies. Since the 1970s, Phase 2a human challenge studies have been conducted to evaluate candidate malaria vaccines, and human challenge studies have been considered as an alternative to Phase 2b field trials for evaluating malaria vaccine efficacy (Sauerwein, Roestenberg, and Moorthy 2011). The WHO recently published a summary of regulatory considerations for human challenge trials for vaccine development (WHO Expert Committee on Biological Standardization 2017b).

9.2 Observational Study Designs

Observational studies are non-randomized, meaning that the experimental vaccine is not randomly assigned. In the absence of randomization, observational studies are subject to measured and unmeasured confounding; thus, the quality of the results will always be viewed as inferior relative to a randomized design. In **Section 9.2.1**, we consider the role of observational studies and those settings where observational study designs are useful. We describe cohort (see **Section 9.2.2**) and case control designs (see **Section 9.2.3**). We further consider studies with field evaluation of vaccines (see **Section 9.2.4.1**), historical controls (see **Section 9.2.4.2**), and quasi-experimental approaches (see **Section 9.2.4.3**).

9.2.1 Role of Observational Studies

Randomized controlled trials are the most reliable and rapid way to identify the relative benefits and risks of vaccines. However, in certain specialized circumstances observational studies may be of value. In settings where an existing licensed vaccine is available with some known efficacy, it may not be ethical to conduct a randomized trial to compare an experimental new vaccine to the existing licensed product. Observational studies may also have value if the goal is to evaluate a reformulation of a licensed vaccine (“bridging” study). Observational studies may also be especially useful in post-licensure phase 4 studies (see **Section 7.1**). Phase 4 studies are essential for assessing long term efficacy of vaccines and to detect less frequent adverse reactions.

Observational studies may be useful in tandem with randomized trials. For example, a randomized trial could expand the population under observation to include participants that fall outside the eligibility requirements for enrolment in the clinical trial. Examples of this approach are discussed in **Section 9.2.2** on cohort studies. Similarly, expanding the population in a parallel cRCT can allow other vaccine quantities, like indirect and overall vaccine effectiveness, to be estimated, as discussed in **Section 9.1.3**.

During PHEs, despite best efforts to explain the rationale for randomization, the randomized trial design may not be acceptable to some communities. Levels of fear and mistrust may be extremely high especially if little is known about a highly lethal disease. Health care systems may be strained to their limits. In such circumstances the logistics of conducting randomized controlled clinical trials during a public health emergency may present enormous hurdles. In this extreme context, sound observational studies are preferred to the alternative of performing no evaluation of vaccine efficacy (Rodrigues and Smith 1999). Nonetheless, it is highly unlikely that observational studies can provide the evidence required to support licensure of an experimental vaccine; as a result, every effort should be made to conduct a randomized trial (see **Section 9.1.1** for a discussion of the role of randomization).

In practice, results of observational studies are easiest to interpret when the effect of the intervention is large enough so as to overshadow random error and bias (Piantadosi 2005:2.3.4). This could occur, for example, if a vaccine is highly efficacious (e.g. vaccine efficacy of 90%) and the study is of sufficient quality to minimize error and bias. Evidence from observational studies is also stronger when the outcome assessment is difficult to bias, say for an incontrovertible event such as all-cause mortality that is reliably ascertained.

Efforts must be made to adjust for potential confounders in both the design and analysis. For example, children with better nutritional status may be less susceptible to disease and may also be more likely to be vaccinated. Other potential confounders could include distance from potential environmental sources of infection, number of people living in households, and whether the individual goes to school or works beyond the home. It is critical to collect data on prognostic factors that lead to individuals receiving vaccination. Reporting of study methods and results should comply with available Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklists (Von et al. 2007). Confounding factors may not be adequately controlled even in high-quality observational studies, which is why such studies represent a lower standard of evidence than randomized studies. Statistical methods, such as propensity scores or marginal structural models used in causal inference, can guide the proper design and analysis of observational studies (Halloran and Struchiner 1991, 1995; Tchetgen and VanderWeele 2012).

9.2.2 Cohort Studies

Description

A cohort study is typically characterized by longitudinal, prospective surveillance of a well-defined population. The approach is similar in structure to a randomized trial except without randomization. For estimating vaccine effects, participants are classified by vaccination status at the beginning of the study period (or at the time of their entry into the cohort), and disease incidence rates are

compared between the two groups. Retrospective cohorts can also be performed if detailed records/database are available on a population such that the study can be expanded beyond a case-control, but we do not consider this setting further. Cohorts can be formed from naturally clustered populations, such as following a series of schools, workers at a few hospitals, or households. Rothman et al (2008:7) summarize the general features of the design, and Orenstein et al. (1985), Halloran et al. (2010:6.2.2), and King et al. (2015) discuss issues specific to vaccination cohorts.

Critical considerations when designing a vaccine cohort study include defining the study population (denominator), the method for ascertaining the endpoint of interest, and the method for ascertaining vaccination status (Orenstein et al. 1985). Vaccination must be completed prior to exposure to the pathogen for the participant to be considered vaccinated in the analysis. Efforts should be made to ensure equal case ascertainment in vaccinated and unvaccinated populations. Total population studies with door to door active case detection are expected to yield the least biased estimate of vaccine effectiveness.

Nested case-control and case-cohort studies are examples of sub-studies implemented within larger cohorts in which all cases are included and the controls represent a subset of the cohort population. These designs may be valuable when the cost of collecting exposure data on all participants who do not develop disease (i.e., controls) is prohibitive. For example, testing all stored blood samples or conducting genetic testing on all disease-free controls may be infeasible. Nested case control and case-cohort studies embedded into cohort studies tend to be of higher quality than standard case-control studies because the source population is very well-defined (Rodrigues and Smith 1999), though a traditional cohort approach is preferred whenever it is feasible for all participants to contribute to the analysis.

ANALYSIS

Vaccine efficacy in a closed cohort can be estimated by cumulative incidence or attack rates. If cohorts are open or dynamic, allowing people to join, leave, or change vaccination status, then it can be based on cases per person-time at risk, incidence rate, or survival analysis methods in which risk set can change over time (M E Halloran et al. 2010:6.2.2). Adjustment for confounders can be performed using many different methods, including a stratified analysis or a regression-based approach with covariates.

Depending on the structure of the cohort, one may be able to estimate indirect, total, and overall vaccine effects (see **Section 8.1**). One example is household studies in which different vaccination coverage levels are observed across households.

SAMPLE SIZE CALCULATION

The sample size for cohort studies depends on the vaccination coverage rate in the population, the assumed vaccine efficacy, the assumed disease attack rate in the controls, and the desired width of the confidence interval for the vaccine effect. Sample size formulae for estimating vaccine efficacy in cohorts are provided in O'Neill (1988). Sample size calculation methods are similar to those used for iRCT designs (see **Section 9.1.2**).

Example

An open prospective cohort study run in two sites in rural Malawi is ongoing to evaluate the post-introduction effectiveness of a 13-valent pneumococcal conjugate vaccine and a monovalent rotavirus vaccine against infant-mortality related outcomes (King et al. 2015). Infants are followed up by field enumerators with a home visit at 4 months and 1 year of age to assess vaccine status, confirm survival, and administer a one-page questionnaire to collect key covariates such as maternal education and household composition. For evaluation of the pneumococcal vaccine, assuming post-neonatal infant mortality of 25/1000 live births, three-dose vaccine coverage of 75%, and 12% loss to follow-up, approximately 35,000 infants are needed to have 80% power to detect vaccine effectiveness of ≥20% against all-cause mortality. Similar calculations were conducted for the rotavirus vaccine. A Cox proportional hazards model is planned to model survival to 1 year and estimate vaccine effectiveness.

Considerations in Outbreaks and PHEs

A vaccination cohort would need to be defined before an outbreak occurs or early in the outbreak so vaccine can be administered prior to most individuals being exposed. Cohorts could be defined based on role, such as a cohort of front-line workers. They could also be defined as households in an area at high risk. Where vaccine coverage is high (e.g., a cohort of health care workers prioritized for vaccination), the cohort population may need to be expanded to include individuals who are ineligible for vaccine or to whom vaccine has not been made available. Individuals ineligible for vaccine will be a less useful comparison group as common exclusion criteria, such as young age or pregnancy, may be associated with different risk of disease.

Cohort studies can be embedded into randomized vaccine trials by expanding the study population beyond randomized participants. For example, an iRCT of front-line workers could be expanded to include monitoring of unvaccinated household contacts. A similar approach has been applied in the past to include the family members of children enrolled in vaccine iRCTs to measure indirect vaccine effects (Trollfors et al. 1998).

An advantage of cohort studies is that they can yield valuable safety and immunogenicity data, especially where baseline samples prior to vaccination and/or prior to infection are collected. Another advantage is that a single study can be used to assess multiple endpoints of interest, whereas case-control studies typically only consider a single case-defining disease endpoint (see **Section 9.2.3**).

Cohort studies may require large samples and be costly to operate. Accurate data on vaccination status and confounding variables are often not available for an entire population or may be expensive to collect, especially in resource-limited settings (Verani et al. 2017b). Cohort studies may take a long time to run, depending on the rarity of the disease. This would be especially true if cohort studies are initiated in inter-epidemic periods when the timing of the next outbreak is unknown. Furthermore, standards of diagnosis may change over time for long studies, which could subject the study to bias (Rodrigues and Smith 1999).

9.2.3 Case-Control Studies

Description

Case-control studies for vaccines are conducted by enrolling disease cases and comparable disease-free controls and comparing vaccination status. Studies can be prospectively integrated into a surveillance program, enrolling cases and controls over time, or be entirely retrospective, using diagnostic or electronic health records. Rothman et al (2008:8) summarize the general features of the case-control design, and Rodrigues and Smith (Rodrigues and Smith 1999), Halloran et al. (2010:8.1), and Verani et al. (2017a, 2017b) summarize the design as used for vaccine studies.

One of the central choices in designing a vaccine case-control study is identifying how to sample controls. The validity of resulting inference depends heavily on the quality of the controls. Controls should have the same risk of exposure to the target pathogen as cases, should be similarly susceptible to the disease before vaccination, should be recruited independently of vaccination status, and should have the same access to medical care and vaccination (Orenstein et al. 1985). Healthy community controls are often selected from the same source populations. A good rule of thumb for selecting controls is that if a control developed the disease of interest, he or she would become a case in the study (Rodrigues and Smith 1999). This is the motivation for the test-negative design, in which individuals testing negative for the pathogen of interest serve as controls. For more details about the test-negative design, see **Box 2**. Similarly, the case-case design is a type of case-control study in which controls are patients attending the same clinics/hospitals but for unrelated aetiologies ("hospital controls").

It is important to define how cases are identified and how vaccination status and covariates are ascertained. Once a target population is clearly defined, efforts should be made to ascertain all cases or a representative sample of cases occurring in that population during the study period (Rodrigues and Smith 1999). This is easiest to achieve for nested case-control or case-cohort designs, which are embedded within a prospective cohort study with a clearly defined population (see **Section 9.2.2**). Cases should be detected using a highly specific test or case definition (Orenstein et al. 1985) as inclusion of false positives biases estimated vaccine efficacy towards the null, especially if the false positive rate varies over time and place (Verani et al. 2017b). A test or case definition with lower sensitivity may be acceptable as long as sensitivity is independent of vaccination status. The drawback of a lower sensitivity test may be failure to recruit enough cases to achieve desired study power (Rodrigues and Smith 1999).

Cases and controls may be matched for key confounders that impact vaccine efficacy, such as age, gender, socioeconomic status, and geography. Matching should be limited to a small number of variables believed to be linked to both vaccination and disease (Verani et al. 2017b). For rare outcomes and/or settings where information on controls is easy to obtain, multiple controls (e.g. 2-4) can be matched to a single case. Pair-matching can add to the cost/complexity of the study, but may strengthen the validity of the results. Stratum-matching, in which multiple cases and controls can fall into any one stratum, may also be considered (Rodrigues and Smith 1999).

The case-control study design does not work well if only a small proportion of the source population is vaccinated because vaccination rates will be low among both cases and controls. By the same logic,

it can be hard to discern vaccine effectiveness if more than 90% are vaccinated. Case-control designs can be especially useful when the outcome is rare (e.g., rare adverse events or rare complications of disease) because the population analysed is enriched with cases, and so large populations do not need to be followed prospectively.

Analysis

In the analysis, estimated vaccine effectiveness is derived as one minus the odds ratio for vaccination. For rare outcomes, the odds ratio closely approximates the risk and rate ratios (Rodrigues and Smith 1999). If the marginal proportions of cases who have been vaccinated and the proportion of the target population who has been vaccinated are known from surveillance, a simple formula can be applied to return a relative risk version of vaccine efficacy (Rodrigues and Smith 1999).

Unmatched case-control studies are typically analysed using the Mantel-Haenszel method, stratifying over all combinations of the confounders, or using unconditional logistic regression. Matched case-control studies can use McNemar's formula or conditional logistic regression. Matching in and of itself does not eliminate confounding, and matching can in certain cases induce confounding if cases and controls are "over-matched" (Brookmeyer, Liang, and Linet 1986). A summary of analysis methods for case-control studies is provided in Pearce (2016) and particularly for vaccine studies in Verani et al. (2017a).

One advantage of matching for vaccine case-control studies is that it may simplify how vaccination status is determined for controls. Investigators typically exclude cases occurring shortly after the last dose of vaccine. In matched case-control studies, the date of disease onset for each case can be used as a "pseudo-date" of disease onset for matched controls. The control is then considered vaccinated only if they received vaccine before the date of disease onset of the matched case (Rodrigues and Smith 1999). When matching is not used, it is unclear how to assign vaccination status to controls who receive vaccine during the study period.

The primary estimand of a case-control study is direct vaccine effectiveness (M E Halloran et al. 2010:6.2.2) (see **Section 8.1**). Since we assume vaccinated and unvaccinated populations interact, we expect the indirect effect to be roughly similar in both groups and so is not measurable in case-control studies (Rodrigues and Smith 1999). Similarly, the total and overall vaccine effects are not estimable.

Case-control studies can assess the effect of multiple exposures, such as different vaccination schedules of number of vaccine doses received. Case-control studies may also be able to assess effect of vaccine waning over time, assuming that cases are regularly observed over time in the population. The analysis is straightforward if information on date of vaccination and date of disease is collected (Rodrigues and Smith 1999).

Sample Size Calculation

Case-control designs typically require smaller sample sizes, and fewer cases, to detect a significant vaccine effect as compared to cohort designs (see **Section 9.2.2**). Sample size for case-control

studies depends on the vaccination coverage rate in controls, the assumed vaccine efficacy, the desired precision of the vaccine effect, and the number of controls per case (if matched). Calculations should further allow for missing data, adjustment for confounding, and the expected prevalence of incomplete vaccination (Verani et al. 2017b). Sample size formulae for estimating vaccine efficacy in case-control studies are provided in O'Neill (1988).

Example

A case-control study was implemented in Sao Paulo, Brazil, to evaluate a serogroup B meningococcal vaccine (de Moraes et al. 1992). The product was licensed, but its efficacy was doubted after it failed to control a meningitis epidemic. The target population for the study was children 3-84 months who lived in Sao Paulo during a period of mass vaccination campaigns. Cases were identified from hospital-based surveillance as children hospitalized with serogroup B meningococcal disease. Four neighbourhood- and age-matched controls were selected for each case. Vaccination cards were examined to determine vaccination history. Conditional logistic regression was used to adjust for confounding variables. Between June 1990 and June 1992, 112 patients and 409 matched controls with confirmed vaccination status were enrolled. Estimated vaccine efficacy was highest in children 48 months or older.

Considerations in Outbreaks and PHEs

To maximize the quality of the study, recruitment of cases and controls would preferably be conducted prospectively rather than retrospectively assessed from medical records. Prospective recruitment imposes greater control over the case definition, assessment of vaccination, quality of matching, and the quality of covariates collected. Retrospective assessment of vaccination status would still typically be required, and this is subject to information and recall bias. Studies may prospectively maintain records of all vaccine recipients in the population in order to easily confirm vaccination status amongst recruited cases and controls. The surveillance system for detecting cases may also need to be strengthened or expanded to limit selection bias. Passive ascertainment of cases could be supplemented with active case finding.

For pathogens causing outbreaks with significant spatiotemporal variability in incidence, matching on variables linked to exposure status will be important. Cases need not have identical exposure to the target pathogen as controls, but other factors could be linked to both exposure and vaccination. For example, if the poorest neighbourhoods targeted for vaccination are also most likely to experience outbreak, it would be important to match controls on neighbourhood to limit confounding (Rodrigues and Smith 1999). Other examples of matching factors might be type of front-line health worker or a co-worker with a shared occupational exposure.

The primary advantage of case-control studies is their relatively low cost because long-term prospective follow-up is not required and the study is enriched with cases, reducing the overall sample size. Case-control studies are especially efficient for rare outcomes. Furthermore, available resources can be directed to a smaller number of cases and controls, allowing for more detailed and

potentially higher quality testing (Orenstein, Bernier, and Hinman 1988). Furthermore, researchers may be more likely to capture serious disease endpoints in the population such as all-cause mortality and hospitalization. In the context of a randomized trial, trial participants exhibiting symptoms would ideally receive swift intervention thereby preventing these types of endpoints (Rodrigues and Smith 1999).

One disadvantage of case-control studies is that prior exposure to the pathogen cannot be measured in participants. As cases are identified following infection, it is not possible to collect any pre-infection samples for immunologic testing. Thus, the vaccine effectiveness estimated may be biased towards the null relative to a randomized trial restricted to a population with no prior evidence of infection (Rodrigues and Smith 1999).

BOX 2: TEST-NEGATIVE DESIGN

In test-negative studies, controls are selected from the pool of people who are tested for the pathogen of interest but test negative (Jackson and Nelson 2013). Test-negative designs attempt to provide control over confounding due to differential case ascertainment, access to care, and health-seeking behaviour. This is achieved by restricting the study population to individuals meeting the clinical case definition who receive testing. Direct vaccine effectiveness is estimated as one minus the odds ratio of vaccination for positive-testing cases versus negative-testing controls.

A central assumption in the test-negative design is that vaccination does not confer cross-protection to other diseases with similar symptoms, as this could bias estimated vaccine effectiveness towards the null (Jackson and Nelson 2013). One way to evaluate this assumption is to check that vaccine coverage in the test-negative controls is similar to vaccine coverage in the underlying study population. If they differ, this could also reflect differences in health-seeking behaviour and may negatively impact study generalizability.

A highly specific test is required to reduce bias in estimated vaccine efficacy (De Serres et al. 2013). The availability of such a test may be a challenge in PHEs where diagnostics are rapidly evolving. It may be necessary to strengthen the laboratory testing infrastructure if an increase in testing is expected. If a more accurate test does not become available until later in the study, properly stored specimens should be retested.

Test-negative designs are an efficient and inexpensive method to estimate direct vaccine effectiveness that can be easily embedded into existing surveillance programs or expanded with the use of remote testing and mobile laboratories. The test-negative design has been extensively applied to measure annual influenza vaccine effectiveness (Jackson and Nelson 2013). The design has been further validated by comparing data from a vaccine iRCT with a test-negative design estimate of vaccine efficacy derived from the trial cohort. High concordance has been found between iRCT vaccine efficacy and test-negative vaccine effectiveness has been found for global influenza vaccine trials (De Serres et al. 2013) and for rotavirus vaccine trials across multiple low-income settings in sub-Saharan Africa and Asia (Schwartz et al. 2016).

9.2.4 Other Observational Study Designs

9.2.4.1 Field Evaluation

Field evaluation of vaccine efficacy is also referred to as the "screening method", "rapid screening", "case-coverage", or "case-population" (Orenstein et al. 1985; Orenstein, Bernier, and Hinman 1988; Rodrigues and Smith 1999). This approach is similar to a case-control approach, but instead of collecting detailed data on a limited control group, surveillance data on the whole population are used. The analysis compares vaccine coverage among cases with the known vaccine coverage in the population. Not all cases need to be ascertained, but cases ascertained should be representative of cases in the population (e.g., a random sample). The primary advantage of this method as compared to a case-control design is that it is cheaper, relying only on surveillance data, but it is considered a crude method for estimating vaccine effectiveness because it allows for no control of confounding (Rodrigues and Smith 1999). It may be used to obtain a preliminary estimate of vaccine effectiveness, hence the terms "rapid" or "screening." The approach should not be relied upon for precise estimates of vaccine efficacy or effectiveness, as would be required to meet regulatory standards. Vaccine effectiveness can be overestimated if the vaccination levels in the community are overestimated or if the proportion of cases with a vaccination history is underestimated. The approach can be used for retrospectively studying outbreaks in a well-defined area that includes both vaccinated persons and unvaccinated persons. Ideally there should be an absence of substantial prior disease activity in the population prior to the outbreak so that population immunity is low.

9.2.4.2 Historical Controls

When a study uses historical controls, the comparator to the treated group is not a concurrent, separate group of patients. The controls could be prior patients from an observational study (e.g. for treatment, prospective natural history study data, medical chart data from clinical care), or a control group from a prior randomized investigational study. The comparison is between two different times, so improvements in disease prevention, changes in diagnostic criteria, changes in population susceptibility, differences in the circulating pathogen, or any other sources of temporal trends can yield uninterpretable results (Piantadosi 2005:2.3.4). Inference is most appropriate when there is limited time between historical data and study data and when the effect is very large. The use of historical controls likely has limited utility in the context of emerging pathogens because spatiotemporal changes in disease transmission may be highly unpredictable. Outbreaks may occur in different areas/populations, and it would be very difficult to disentangle the effect of vaccine from these sources of variability. A useful discussion on the role of historical controls in the context of AIDS clinical trials is available in Byar et al. (1990).

9.2.4.3 Quasi-Experimental Design

Quasi-experimental studies are observational (non-randomized) studies designed to resemble a randomized trial, but they can be applied in settings where a randomized trial is infeasible or otherwise unacceptable. These designs typically compare groups across some boundary, assuming that the differences between the groups are minor and that the groups are otherwise comparable.

Interrupted time series analysis estimates the effect of an intervention that is introduced at a particular time, and so the boundary is temporal (before versus after). Analytical methods for time series can be used to attempt to adjust for secular trends. The approach shares some of the limitations of historical controls (see **Section 9.2.4.2**), though there is expected to be no gap in data collection, consistency in the population, and no change in the methods used for case ascertainment. The analysis can be conducted using population-level data, and provide an estimate of overall effectiveness of the program. Interrupted time series analysis has been used to evaluate the population-effects of rotavirus (do Carmo et al. 2011) and pneumococcal vaccination programs (Grijalva et al. 2007).

Regression discontinuity methods similarly estimate the effect of an intervention across a temporal boundary or other type of boundary. For example, only a certain area may receive vaccine, but neighbouring areas share similar risk factors. Alternatively, we could compare incidence across an age boundary (e.g. comparing 17 year olds with 18 year olds for a vaccine licensed for 18+), as individuals age into eligibility for a vaccine. Regression methods applied to population- or individual-level data are used to estimate differences across this boundary that are likely attributable to the vaccine. Regression discontinuity methods were used to estimate the effect of a mass polio vaccination campaign in Bangladesh (Helleringer, Asuming, and Abdelwahab 2016). Children in the target age range during the campaign were compared to children who were slightly too old to have received vaccination during the campaign.

10 Additional Design and Conduct Considerations

In this section, we describe other design and conduct considerations, primarily those that relate to the data analysis. In **Section 10.1** we discuss data monitoring strategies, focusing on the particular setting of waning outbreaks in **Section 10.1.1**. In **Section 10.2** we consider approaches for accumulating evidence across outbreaks, which may be necessary for pathogens that cause small or unpredictably sized outbreaks in humans. In **Section 10.3** we describe strategies for preventing and handling missingness in the data. In **Section 10.4**, we discuss the use of simulations to assist in trial design and interpretation.

10.1 Data Monitoring

In outbreaks of emerging pathogens, there is an imperative to provide timely access to reliable insights about effective interventions. The interest in timeliness should be balanced with the need to avoid risks for pre-judgments occurring when access is provided to emerging data from clinical trials that may be more misleading than insightful.

To achieve such balance, an independent data safety monitoring board (DMSB) should be in place. The mission of the DSMB is to safeguard the interests of study participants and to enhance the integrity and credibility of the clinical trial (Ellenberg, Fleming, and DeMets 2002). Some principles of fundamental importance to achieving their mission is that the DSMB should have sole access to emerging results on relative efficacy and safety of interventions in the trial, should have multidisciplinary representation with members having proper experience in the DSMB process, and should be independent with freedom from apparent significant financial, professional or regulatory conflicts-of-interest.

To guide the DSMB in their interpretation of interim data, the design of the clinical trial should include specification of data monitoring boundaries. These guidelines help the DSMB differentiate between reliable evidence about the effects of the experimental interventions versus fluctuations occurring over time that are consistent with random variability.

When the primary endpoint of a clinical trial is a measure of major morbidity or mortality, it is particularly important to establish a data monitoring boundary to provide guidance about whether interim data are sufficiently favourable to justify termination of the trial with a conclusion of benefit, while controlling the trial's overall Type I error rate. A boundary also can be established to provide guidance about whether interim data are sufficiently unfavourable to justify a conclusion of lack of benefit or "futility", in a manner that avoids eroding the statistical power of the trial. It is recommended that a data monitoring strategy be specified for key secondary endpoints.

Group sequential guidelines, such as an O'Brien-Fleming boundary, provide a widely implemented approach, (O'Brien and Fleming 1979; Pocock 1977). By implementing these boundaries through an alpha spending approach (Lan and DeMets 1983), there is important flexibility in the number and timing of the interim analyses (see Chapter 8 of (Ellenberg et al. 2002)). Adapted approaches are available for monitoring of cRCTs (see **Section 9.1.3**) (Zou, Donner, and Klar 2005).

If a clinical trial is terminated early, prompt sharing of study data is encouraged (Lane, Marston, and Fauci 2016). There should be some plan in place in the protocol for next steps, which may include vaccinating all eligible, consenting, unvaccinated participants. These participants should then be followed for safety outcomes since the product is unlicensed.

A double blind, individually randomized trial of a heptavalent pneumococcal conjugate vaccine was conducted in California (Black et al. 2000). Over 37,000 infants were randomized 1:1 to the conjugate vaccine or meningococcus type C conjugate (comparator group). The primary outcome was invasive disease caused by vaccine serotype. At the interim analysis, all cases of invasive disease had occurred in the comparator group, yielding an estimated vaccine efficacy of 100%. Termination of the trial was recommended by the Study Advisory Group (Black et al. 2000).

Data monitoring is distinguished from the broader class of adaptive design methods available for clinical trials. Adaptive designs use unblinded study data to make decisions. Changes to the protocol that affect both treatment groups equally and are made blinded to study data are generally not considered adaptive (Byar 1990), such as using updated standard of care procedures for case isolation. One example of adaptive design in vaccine trials is modifying the sample size after an interim analysis for settings where the incidence rate is unknown (Li, Chan, and Anderson 2012). One key limitation of adaptive design methods is that a subset of the data is used to both generate and then confirm study hypotheses, which can make results less interpretable. Adaptive procedures

must be pre-specified in the protocol, which can have the unintended effect of limiting flexibility. Guidance on adaptive design methods is available elsewhere (Food and Drug Administration 2010).

10.1.1 Planning for Settings where Epidemic Wanes

Adequate guidance does not exist on how to conduct data monitoring in settings where an epidemic wanes such that transmission among humans declines to extremely low levels or stops entirely. In this scenario, a study may not be able to accrue further evidence to directly evaluate protective vaccine efficacy. If this occurs, the study protocol should clarify how data will be analysed as the full sample size has not been reached.

A waning epidemic could trigger study closure, study pause, study continuation to collect additional data, or a formal evaluation to decide on next steps. Study closure would trigger an analysis of final data, and it should be pre-specified how remaining alpha will be spent at this time. In settings where the epidemic wanes relatively early after the start of the trial, it may be necessary to specify a minimum sample size or number of events for the data analysis to be considered reliable. Investigators could pause the study until the next outbreak occurs in the study area; approaches for combining evidence across outbreaks are discussed further in **Section 10.2**. Investigators may also choose to keep the study open to collect additional safety and immunogenicity data. Keeping the study open would also be desirable in case there is an unexpected surge in transmission. Finally, a waning epidemic could trigger a formal evaluation to decide whether the study should remain open, paused, or closed. This evaluation could rely on transmission modelling to assess the probability of future cases in the current outbreak or future outbreaks in the study area. This type of modelling has been used to inform likely case accrual for Ebola vaccine trials (Camacho et al. 2015), and it is discussed further in **Section 10.4**.

In practice, it is necessary to establish how a waning epidemic is defined for the purposes of the study. Decisions about waning could be pre-specified or could be made using data fully independent of (blinded to) the study data. Examples of blinded approaches include waiting until the outbreak is declared over by the international health community or when no transmission has occurred for at least two serial intervals in the study area. Alternatively, the overall (combined) incidence rate for participants in the study could be examined. Examples of unblinded approaches include measuring the unblinded incidence rates for the two arms of the study.

10.2 Accumulating Evidence about Interventions across Outbreaks

For emerging infectious diseases, it may be difficult to accumulate enough evidence to reliably ascertain the efficacy of an intervention from a single outbreak. It is also generally not possible to predict the size and duration of an outbreak or the extent of its spread; consequently one cannot provide an accurate prediction of the number of participants that will be available for clinical trials at the time of their implementation. As a result, it may be necessary to proactively plan to combine information across multiple outbreaks (or trials within the same outbreak) to evaluate the efficacy of an intervention. Such an approach might be desirable for diseases like Lassa fever.

In this setting, trialists may consider a few different approaches for accumulating evidence across outbreaks. The first approach is a conventional, sequential clinical trial in which the protocol can be paused or suspended between outbreaks. Interim analyses would take place at the end of each outbreak, at which recommendation for stopping would be based on demonstration of efficacy or futility. Alpha spending, as defined by the fraction of required information for the master protocol achieved by the end of each outbreak. The second approach allows for separate trials at each outbreak. But if any of these trials did not reach the targeted amount of information, the study DSMB could choose to recommend against release of study results publicly. This choice would allow for the possibility of the study data to be combined with that of future trials-- of the same product and with similar design. If in a later outbreak, the new study does not accrue the required information, the sponsors/conductors of that study could make a decision to pool the data from both trials (and potentially from additional studies as well) in a final analysis. A central requirement is that that no one involved in decision-making has any access to unblinded data from either study. The third approach is a prospective or pre-planned meta-analysis of several possibly underpowered studies. A prospective approach encourages standardization across trials, such as alignment of trial endpoints, and thereby can improve interpretability than. Meta-analyses or overview studies cannot serve as substitutes for well-conducted clinical trials as they do not have a mechanism for preserving type I error, but they can provide valuable summaries of intervention effects.

Results from studies across multiple outbreaks may also be integrated using Bayesian approaches. The first study uses a non-informative prior, and the posterior distribution of the parameter of interest is used as the prior in the next study, and so on. As frequentist methods predominate in clinical trials, it would be important to predefine a data monitoring strategy for this Bayesian approach that is acceptable to regulators.

10.3 Missingness in Data

"The reliability and interpretability of results from clinical trials can be substantially reduced by missing data" (Fleming 2011; National Research Council of the National Academies 2010). This is particularly apparent when participants who are lost to follow-up have different risk factors for disease. Such missingness induces risk for substantial bias.

Commonly used approaches to 'treat' missingness, such as conducting complete case analyses or use of last-observation-carried-forward methods, rely on strong assumptions that usually are invalid. Hence, to more effectively protect the integrity of research, there should be proper focus on approaches to prevent rather than simply to treat missingness.

Achieving low levels of missing data requires active rather than passive approaches. These approaches should include several aspects (Fleming 2011; National Research Council of the National Academies 2010).

First, protocols should properly distinguish between withdrawals from the randomized intervention (i.e., non-adherence) versus withdrawals from follow-up (i.e., non-retention). While there are many valid reasons to be non-adherent, the only valid reasons for non-retention are death or 'withdrawal of consent'. The term 'withdrawal of consent' should be properly applied. Investigators should be educated that this term should be used, not simply because the participant no longer wishes to

receive their randomized intervention or to return for follow-up evaluations, but rather because they no longer authorize study investigators to make efforts to obtain their outcome data. 'Withdrawal of consent' should be initiated by the participant and ideally would be done in writing.

Second, during the informed consent process, while participants should be informed about their right to withdraw consent at any time, they also have the right to be informed that doing so would induce risk of substantive bias that would compromise trial integrity, in turn diminishing the scientific value of the altruistic contributions of the study participants.

Third, studies should engage investigators who are committed to following each surviving participant until the capture of all trial outcomes, even if the participant has discontinued their randomized intervention or initiated other interventions, in order to enable the conduct of intention-to-treat (ITT) analyses (see **Section 8.2.2**). Only ITT analyses preserve the integrity of randomization and properly evaluate an experimental intervention in the context of an intervention strategy. An ITT analysis requires that outcomes are ascertained in all participants. Where missing data persists, sensitivity analyses are recommended to explore the effect of departures from strong assumptions (White et al. 2011).

Fourth, protocols should specify performance standards for achieving high quality of trial conduct, including high levels of data capture; in turn, creative and effective approaches should be implemented during enrollment and follow-up to enhance achieving pre-specified targeted levels of retention; finally, a 'peer review' oversight process should be in place to ensure the achievement of the performance standards, including achieving targeted levels of data capture.

10.4 Simulating Trials for Planning and Interpretation

Typical sample size calculations for clinical trials use closed-form expressions to determine the required number of endpoints or total sample size to achieve a desired level of power. These expressions may not adequately capture the complex non-linear dynamics of infectious diseases, including the indirect effects of vaccination (see **Section 8.1**) and interactions between vaccine and other disease containment measures. There may also be important sources of heterogeneity in hosts, pathogens, and exposure, as well as logistical complexities in implementing the trial that alter the timing of the trial. It may be difficult to incorporate these important factors into the design process (Halloran et al. 2017).

Mathematical and computational simulations can play a key role in the design and interpretation of vaccine trials. Simulations allow researchers to capture the complex dynamics described above, to consider other key outcomes such as trial duration, cost, and overall impact on the epidemic, and to interpret unexpected trial results. Simulations can vary in complexity, from simple dynamics to full calibration to the trial region. Stochastic simulations can more accurately reproduce population-level effects than deterministic models. Additional recommendations about the structure of simulations for infectious disease trials are described in Halloran et al. (2017).

11 Operational and Implementation Issues

Studies conducted during outbreaks and PHEs face key logistical challenges. These challenges influence the design choices which will maximize the study's likelihood of success.

Countries experiencing the outbreak may or may not have the clinical and laboratory facilities needed to conduct the study. A lack of nearby laboratory facilities at the appropriate biosafety level (BSL) may be problematic if biological samples are reliant on adequate cold chain from the study site to a distant lab. Collection and movement of samples may be difficult in some contexts, such as where the roads are poor, or where disruption/unrest is occurring in some regions. Studies may take advantage of new technology for collecting blood samples and maintaining a cold chain. The Ebola vaccine trials in Guinea and Sierra Leone used special transport canisters for vaccine delivery, although these are not appropriate for long-term storage. Mobile labs may allow some testing to be conducted on-site. In general, though, these barriers may slow down vaccine delivery and confirmation of cases.

Local study staff may be unfamiliar with certain types of technology used for data collection. Staff may be quickly trained in the operation of tablets, which are useful because they are portable and allow remote data collection that can be uploaded when internet access becomes available. Where traditional paper collection is used, there may be a delay because data must be later entered into computers at a centralized location by other staff. Teleconferences may also be very challenging to conduct with people in the field. Responsive designs typically require rapid availability of information for informing the study (see **Section 8.3.3**); these designs may not be feasible or may be less effective if there are significant delays in vaccine delivery, specimen collection and testing, and the availability of results to study staff. Similarly, designs that rely on complex schemes for randomization (e.g., adaptive randomization) will be difficult to implement without real-time electronic communication with study headquarters.

Typically studies require a high degree of up-front training and planning. Stepped or phased rollout of the study may improve feasibility. With the exception of stepped wedge cRCTs (see **Section 9.1.4**), study sites may be added when they become available (e.g., after setting up the cold chain and training staff). If a sufficient supply of the vaccine is not available at the start of the trial, the sample size may be gradually increased over time. In the PREVAIL I trial Ebola vaccination trial, researchers implemented caps on the number of participants to be vaccinated per day due to the daily limit of vaccine syringes prepared; as they were not able to consent and vaccinate all interested participants on a given day, the team provided reservations to additional participants in the order that they volunteered to participate (Kennedy et al. 2016).

Limited supply of vaccine and/or eligible at-risk participants is also relevant for the setting of multiple, simultaneous vaccine trials. It is recommended that multiple trials occur simultaneously, if resources permit. This allows trials to cover different areas and to include different groups of participants in different designs. Where vaccine supply is limited or transmission is only occurring in a limited geographic area, there is an unfortunate risk that two or more trials may compete/neutralize each other to enrol participants. Independent trials are encouraged to communicate and collaborate as much as possible so that the methodology is aligned enough so that results are conducive to a meta-analysis (see **Section 10.2**).

Country ownership is critical for the success of vaccine studies in developing countries. Local researchers are best equipped to assess limitations of their laboratory, clinical, and surveillance

infrastructure. In some cases, it may be necessary to enhance existing infrastructure or to develop a parallel infrastructure for the study. Study staff may include a mix of national and international staff. Ideally all study staff will be familiar with GCP. While international staff may have more prior experience with the implementation of clinical studies, there may be poor continuity with staff rotating in and out of the country.

Community engagement is critical to establish community acceptability of control arms, placebos, and blinding. The WHO has prepared guidelines on good participatory practice (GPP) and community engagement in this setting (WHO 2016c) Acceptability may impact whether trials are individually or cluster randomized, blinded or unblinded, and use a placebo or other comparator. Randomized designs that rely on a placebo or other control comparator may be more easily justified to participants and local communities when a concrete commitment of post-trial resources is made (e.g., either rollout of vaccine if found efficacious, or other permanent infrastructure created during trial).

Unique cultural considerations may impact the study design. For example, in West Africa, blood draws are not a routine or familiar part of medical care. Thus, the process was unfamiliar to local staff and participants in Ebola vaccine trials. These considerations may affect the choice of endpoints collected in the study.

Safety issues may also impact the study design. For example, for mobile teams taking blood samples for highly infectious pathogens such as Ebola, there is a serious risk of accidental exposure. Thus, collection of blood samples may not be feasible in the less than ideal conditions in the field. Availability of appropriate clinical facilities may also impact the study design. For Ebola vaccine trials, it was necessary to consider how to advise people who may develop fever as this could be a side effect of vaccine or an early warning sign of Ebola. While citizens with fever were advised to undergo Ebola testing, this may put uninfected individuals at increased risk of exposure by visiting testing facilities (Nason 2016). The STRIVE vaccine trial slightly modified the case definition to allow a short 48 hour delay in determining whether a participant with fever had suspected Ebola, though people with Ebola exposure or Ebola symptoms inconsistent with vaccination were immediately treated (Widdowson et al. 2016).

12 Outbreak- and Vaccine-Specific Considerations

Finally, we summarize how characteristics of the pathogen/outbreak (see **Section 12.1**) and of the particular vaccine candidate(s) being evaluated (see **Section 12.2**) impact the recommended study design. The ability to incorporate such context-specific information into the design will depend upon when that information becomes available. Thus, we also consider the impact of uncertainty on study design.

12.1 Impact of Outbreak Characteristics

In this section we describe how key characteristics of the disease, the local infrastructure, and uncertainty about future spatiotemporal incidence may impact the desired study design. Available

information to guide study design may be limited if it is the first major outbreak in a human population, with prior outbreaks being small or poorly understood.

12.1.1 Natural History of Disease

Characteristics of the pathogen which influence the choice of study include the modes of transmission, the incubation and infectious periods, the pathogenicity, the range of severe complications, and groups at highest risk of severe complications.

The first key consideration is the pathogen's mode of transmission, which in turn influences which individuals are at highest risk of infection and how the study population might be defined (see **Section 8.3**). The study population should reflect the pathogen's primary route of transmission. Vector-borne diseases might target areas with high densities of the vector and consider a general population or responsively defined population strategy with spatial/geographic areas serving as study sites. For diseases that spread via direct physical contact, including sexual transmission, the population could be defined to capture contact networks either prospectively before transmission begins or in response to a confirmed case. Where contact networks are difficult to identify, the study could capture communities or spatial/geographic areas. For sexually transmitted diseases, the trial could identify a high risk population based on behavioural risk factors. For pathogens where nosocomial transmission is a key driver, hospitals serve as natural study sites. For pathogens whose outbreaks are caused by multiple spillover events from an animal reservoir, studies could identify groups at high risk of infection based on exposure to animals, which could include employment as mine workers or abattoir workers, or exposure to sick individuals, such as medical front-line workers. These strategies use our understanding about the pathogen's mode of transmission to identify individuals or networks at high risk of infection.

Next, the incubation and infectious periods of the pathogen should be considered. The incubation period is the time from infection to symptom onset. The infectious period is the time during which the infected individual is infectious to others. Responsive designs (see **Section 8.3.3**) are more feasible for studies of diseases with slower natural histories, including longer incubation periods, because they provide more time for investigators to conduct the necessary activities to implement responsive vaccination. For some pathogens, infectiousness begins prior to symptom onset. Pathogens for which a high proportion of transmission occur prior to symptoms are difficult to contain using standard control measures such as case isolation, contact tracing, and quarantine (Fraser et al. 2004), so these pathogens may not be well-suited for responsive designs using contact tracing. Alternatively, a responsive design in which the selection of an entire study site is triggered by one (or several) confirmed cases in the area may still work because a larger network beyond immediate contacts is included.

The pathogenicity, which is the proportion of infections resulting in symptoms, may impact the choice of primary endpoint (see **Section 8.2.1**). Though laboratory confirmed clinical disease is typically the endpoint of public health interest, if pathogenicity is very low, the required sample size will dramatically increase. In limited settings, the use of infection with or without symptoms may be justified as the primary endpoint to improve feasibility. Detecting asymptomatic infections will

require additional specimen sampling and laboratory testing, which in turn depends on the availability of reliable and validated diagnostic assays.

Pathogen virulence further influences the study design. Cases of severe disease, especially those that require hospitalization or result in mortality, are generally easier to detect and will be a more reliably documented study endpoint than any symptoms (see **Section 8.2**). Passive surveillance, which only includes cases seeking medical attention, may be considered when disease is inevitably severe and highly likely to be identified by physicians or public health authorities. Cases with mild or moderate symptoms are more likely to go undetected unless active surveillance is used for case finding. If there are special populations who are at especially high risk of severe disease, examining their outcomes may be a scientific priority (see **Section 8.3.4**). The severity of the disease may also influence the choice of comparator arm (e.g. placebo, active control, delayed vaccination) (see **Section 8.4**).

The distinctiveness of the clinical syndrome impacts the study design. Pathogens may result in symptoms that are similar to other endemic infections (e.g. dengue and Zika) or conditions. Co-circulating pathogens may reduce our ability to recognize cases without an adequately specific diagnostic assay, and a primary endpoint without laboratory confirmation would be highly discouraged (see **Section 8.2.1**). Malnutrition and other local factors could also impact vaccine immunogenicity. In some cases, co-circulating pathogens could even influence vaccine efficacy or other immunological endpoints (e.g. HIV).

The information described above would be obtained from the epidemiological and medical literature. For some emerging pathogens, especially those that have caused only limited outbreaks in humans, data may be scarce or unreliable. In these settings, pilot studies may be especially valuable. Investigators may also plan sub-studies within the vaccine study to collect additional epidemiologically important information.

12.1.2 Local Infrastructure

The local context where the study will be conducted, including the available surveillance, laboratory, and clinical infrastructure will also impact its design.

Investigators must consider the existing medical infrastructure and disease surveillance system when defining the study endpoints (see **Section 8.2**) and sample size. Investigators may choose to align trial endpoints with case definitions used in outbreak control to leverage ongoing surveillance. For pathogens with high morbidity and mortality, surveillance may be more complete, but vaccine studies for pathogens causing less severe disease would typically require additional active surveillance. The type of active surveillance will depend on how frequently symptomatic individuals seek care, if they are regularly tested for infection, and how long it takes for this information to be recorded and reported into the system. Individuals in resource-limited settings with few doctors or no clinics nearby may be less likely to seek care. If case ascertainment and surveillance is poor, it may be hard to use the existing system to inform responsive designs (see **Section 8.3.3**).

Ideally study sites, such as hospitals, would have experience with clinical research, but this may not be the case in resource-limited settings where outbreaks of emerging pathogens occur. Clinical and

laboratory capacity may be poor, impacting the feasibility of frequent blood draws, transportation and storage of samples, and the types of diagnostic assays used; operational considerations are discussed further in **Section 11**. It may be necessary to strengthen/supplement local infrastructure, which will require expenditure of more time and resources before the study can begin. Studies may need to be smaller or use phased rollout to avoid overburdening the system.

For outbreaks of emerging pathogens, other control measures such as case isolation, contact tracing, and quarantine will likely be in place to reduce transmission. The vaccine study should not impede or interfere with these efforts. It is important to consider these other control measures when designing the trial. Control measures that are increasingly effective over time at getting infected individuals into care sooner, may result in important temporal trends in disease severity as cases access treatment earlier. In this case, disease severity may not a desirable primary endpoint (see **Section 8.2.1.4**).

Investigators should also closely consider the relevant aspects of the sociocultural context that may affect the study design, such as a history of distrust of local, national, or international authorities, or any past experience with clinical research studies. Operational and implementation issues are discussed further in **Section 11**.

12.1.3 Current and Future Spatiotemporal Incidence

A major challenge in implementing vaccine studies for emerging pathogens is the unpredictable spatiotemporal incidence of future cases. Investigators must assess the extent of the current outbreak, including magnitude and geographical spread. Small outbreaks may continue to grow or could be quickly controlled. Large outbreaks may be harder to control, but it is possible that the period of peak incidence has already passed and the rate of new cases may begin to decline.

Mathematical models can be used to inform our understanding of future spread, but they rely on high quality input data on both the natural history of disease and the current status of the outbreak as measured by surveillance. Even with high quality data, there may be great uncertainty at the time horizons considered for vaccine study planning (e.g. one to two years in the future). Studies should consider the possibility that the outbreak will end before enough data have accrued; this setting is further described in **Section 10.1.1**. Rapid planning and implementation of studies is of critical importance when the future of the outbreak is uncertain. If the pathogen causes many small outbreaks, it may not be possible to accrue enough evidence from a single outbreak to reach a conclusion; recommendations for combining information from studies in multiple outbreaks is described in **Section 10.2**.

The duration of the outbreak may be difficult to predict, and in some cases, the location of the outbreak is also difficult to predict. Diseases may have high spatial heterogeneity, especially if small outbreaks are caused by independent spillover events from an animal reservoir. In this case, investigators should examine both the geographical range of the current outbreak and the range of historical outbreaks or areas where the reservoir population is known to be infected with the pathogen. If this is a wide area, a general population study may not be realistic as the overall incidence may be too low for a feasible sample size. A responsive design in which study sites are added when transmission spreads to a new area may be preferred (see **Section 8.3.3**), but this

approach may require enhanced surveillance in these regions to rapidly detect new cases and initiate recruitment and vaccination.

12.2 Impact of Vaccine Candidate Characteristics

In this section we describe how key vaccine characteristics may impact the desired study design. Information about the vaccine candidate(s) may be limited if Phase 1 and 2 trials have not been completed at the time of planning. Completing Phase 1 and 2 trials during inter-epidemic periods is recommended to accelerate the process and improve the quality of Phase 3 trials.

12.2.1 Vaccine Efficacy and Immunogenicity

The predicted vaccine efficacy is important for sample size and design. There may be substantial uncertainty in estimated effect size, in which case a conservative (lower) level of vaccine efficacy, still considered to be acceptable, should be assumed for planning. A sequential design to allow for interim data analysis with plans to terminate the trial early if the vaccine is much more efficacious than planned (see **Section 10.1**) can be implemented.

There may be heterogeneity in the vaccine effect by individual characteristics. Investigators should consider covariates or other prognostic factors that are likely to influence the response to vaccination in a systematic way, such as age or prior exposure to disease. Potential covariates of interest may be identified from earlier phase vaccine trials, familiarity with the vaccine platform, or experiences with related pathogens. Investigators should ensure that the study population captures participants with a range of values for the covariates of interest. For immunological covariates such as prior exposure to disease, it may be necessary to collect and store baseline samples from participants. Covariates that are believed to be strong predictors of outcome should be considered as stratification factors in the randomization plan.

The time it takes for vaccinated individuals to develop an immune response, sometimes called the ramp-up period, also impacts the study design. Vaccines may require multiple doses spaced out over weeks or months, or booster vaccinations to achieve maximum efficacy. Slow-acting vaccines may not be well-suited for responsive designs because participants may be exposed to the pathogen before their immune system has developed a protective response (see **Section 8.3.3**); vaccine efficacy will be under-estimated if individuals are not fully protected when they are exposed. If a vaccine has the potential for post-exposure effectiveness, this effect could be evaluated using a responsive design, targeting individuals already exposed to infectious cases.

Where possible, it is recommended that immunological variables be measured in participants to assess correlates of risk (see **Sections 7.2** and **8.2.3.1**). Results from early phase trials as well as an understanding of the vaccine platform help guide the selection of immunological variables. For example, vaccine candidates using the same backbone as other vaccines will have more information about biomarkers on which to focus (Jacobsen et al. 2016). Investigators may also consider experiences from related pathogens for which effective vaccines are available.

12.2.2 Vaccine Safety

The expected safety profile of the vaccine has a further impact on the study design. Where available, investigators would consider safety results from early phase trials when designing the safety monitoring plan for the vaccine efficacy study (see **Section 8.2.4**). If early phase trials are implemented during the same outbreak, complete results may not be available at the time of study planning. For Ebola vaccine trials, in some cases Phase 2 enrolment began in the absence of published Phase 1 data (Nason 2016). These trials may also be so small that they only provide data on common adverse events. Thus, rare and serious adverse events may only be detectable in Phase 3 studies in these settings. Thus, wherever possible, Phase 1 and 2 studies should be completed in inter-epidemic periods.

Side effects from vaccination may include standard effects such as pain and/or swelling at injection site and transient fever. For diseases in which side effects like fever resemble early stages of clinical illness, it will be necessary to define a strategy for appropriately testing and treating these cases. For example, uninfected participants with fever following Ebola vaccination may be put at increased risk if sent to Ebola treatment units, where they could be exposed to highly infectious cases.

For candidate vaccines with higher safety concerns, the study population may be restricted to individuals at highest risk of infection or highest risk of severe disease since the anticipated risks may outweigh benefits in the general population (see **Section 8.3.2**).

12.2.3 Operational Considerations

For vaccines developed during outbreaks in which the pace of manufacturing is accelerated, it is necessary to consider the impact of the timing of the vaccine production process. For example, only a limited number of doses may be available, or doses may only become available over time. There may be unexpected delays in the manufacturing process; this is especially relevant for vaccines made on new platforms.

In settings where the number of doses of vaccine is limited, a high risk population (see **Section 8.3.2**) or responsively defined population (see **Section 8.3.3**) strategy may be preferred to minimize the required sample size. Similarly, individual randomization (see **Section 9.1**) may be preferred over parallel cluster randomization trials (see **Section 9.1.3**) to minimize the required sample size.

If limited doses of vaccine are available at the start of the trial and more only become available over time, investigators may prefer designs with naturally phased roll-out, such as studies with responsively defined populations in which sites are added as transmission spreads to new areas (see **Section 8.3.3**) or stepped wedge trials (see **Section 9.1.4**).

Finally, the specific requirements for storage, preparation, and administration of vaccine candidates will dictate operational considerations. Staff may require additional training to administer the vaccine. Certain vaccines may be difficult to blind to participants and vaccinators (e.g. lyophilized liquid vaccines) (see **Section 8.4.1**). Vaccines may need to be maintained at a cold chain that impacts study logistics. Vaccines that are more stable and have minimal storage requirements are preferred in resource-limited settings, but a candidate with this profile may not be available. Vaccines with a short shelf life may not be well-suited for trials with long duration, such as stepped wedge trials (see **Section 9.1.4**) or trials spanning multiple outbreaks (see **Section 10.2**).

13 DISSEMINATION OF THE GUIDANCE DOCUMENT

The overall goal of this guidance document is to improve vaccine evaluation in PHEs. Dissemination and implementation of the recommendations are to be considered by all actors participating in the design, implementation and analysis of vaccine trials at the international, national and local levels. The recommendation made in this guidance document will be disseminated through WHO regional and country offices, ministries of health, professional organizations, WHO collaborating centres, other United Nations (UN) agencies, and nongovernmental organizations, among others. This guidance will also be available on the WHO R&D Blueprint and Initiative for Vaccine Research websites. To increase awareness of the guideline, a manuscript will be published in a peer-reviewed journal. The R&D Blueprint team will support the implementation of the guidance document through workshops in affected countries and the use of existing platform such as the African Vaccine Regulatory Framework (AVAREF).

REFERENCES

Anon. n.d. "WHO | Blueprint Accelerating R&D Processes." *WHO*.

Baird, Sarah, J.Aislinn Bohren, Craig McIntosh, and Berk Özler. 2016. "Optimal Design of Experiments in the Presence of Interference." *ArXiv* December. Retrieved (http://arxiv.org/abs/1507.00803).

Barker, AD et al. 2009. "I-SPY 2: An Adaptive Breast Cancer Trial Design in the Setting of Neoadjuvant Chemotherapy." *Clinical Pharmacology & Therapeutics* 86(1):97–100. Retrieved (http://doi.wiley.com/10.1038/clpt.2009.68).

Bellan, Steven E. et al. 2015. "Statistical Power and Validity of Ebola Vaccine Trials in Sierra Leone: A Simulation Study of Trial Design and Analysis." *The Lancet Infectious Diseases* 15(6):703–10.

Belshe, Robert B. et al. 2007. "Live Attenuated versus Inactivated Influenza Vaccine in Infants and Young Children." *New England Journal of Medicine* 356(7):685–96. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/17301299).

Benn, C. S. et al. 2010. "Vitamin A Supplementation and BCG Vaccination at Birth in Low Birthweight Neonates: Two by Two Factorial Randomised Controlled Trial." *Bmj* 340(mar09 1):c1101–c1101. Retrieved (http://www.bmj.com/cgi/doi/10.1136/bmj.c1101).

Bhattacharya, Sujit K. et al. 2013. "5 Year Efficacy of a Bivalent Killed Whole-Cell Oral Cholera Vaccine in Kolkata, India: A Cluster-Randomised, Double-Blind, Placebo-Controlled Trial." *The Lancet Infectious Diseases* 13(12):1050–56. Retrieved (http://dx.doi.org/10.1016/S1473-3099(13)70273-1).

Black, Steven et al. 2000. "Efficacy, Safety and Immunogenicity of Heptavalent Pneumococcal Conjugate Vaccine in Children." *Pediatric Infectious Disease Journal* 19(3):187–95.

Boren, David et al. 2014. "Stochastic Variation in Network Epidemic Models: Implications for the Design of Community Level HIV Prevention Trials." *Statistics in Medicine* 33(22):3894–3904.

Borm, George F., R. J. F. Melis, Steven Teerenstra, and Petronella G. Peer. 2005. "Pseudo Cluster

Randomization: A Treatment Allocation Method to Minimize Contamination and Selection Bias." *Statistics in Medicine* 24(23):3535–47.

Brookmeyer, R., KY Liang, and M. Linet. 1986. "Matched Case-Control Designs and Overmatched Analyses." *American Journal of Epidemiology* 124(4):693–701.

Byar, David. 1990. "Design Consideration for AIDS Trials." *Journal of Acquired Immune Deficiency Syndromes* 3(2):16–19.

Byar, David P. et al. 1990. "Design Considerations for AIDS Trials." *The New England Journal of Medicine* 323(19):1343–48.

Camacho, Anton et al. 2015. "Estimating the Probability of Demonstrating Vaccine Efficacy in the Declining Ebola Epidemic: A Bayesian Modelling Approach." *BMJ Open* 5(12):e009346.

Capeding, Maria Rosario et al. 2014. "Clinical Efficacy and Safety of a Novel Tetravalent Dengue Vaccine in Healthy Children in Asia: A Phase 3, Randomised, Observer-Masked, Placebo-Controlled Trial." *The Lancet* 384(9951):1358–65. Retrieved (http://linkinghub.elsevier.com/retrieve/pii/S0140673614610606).

do Carmo, Greice Madeleine Ikeda et al. 2011. "Decline in Diarrhea Mortality and Admissions after Routine Childhood Rotavirus Immunization in Brazil: A Time-Series Analysis." *PLoS Medicine* 8(4):11.

Carter, Richard. 2001. "Transmission Blocking Malaria Vaccines." *Vaccine* 19(17–19):2309–14.

Chang, MN, HA Guess, and JF Heyse. 1994. "Reduction in Burden of Illness: A New Efficacy Measure for Prevention Trials." *Statistics in Medicine* 13:1807–14.

Clemens, John, Ruth Brenner, Malla Rao, Nebiat Tafari, and Charles Lowe. 1996. "Evaluating New Vaccines for Developing Countries: Efficacy or Effectiveness?" *Journal of the American Medical Association* 275(5):390–97.

Clements-Mann, ML. 1998. "Lessons for AIDS Vaccine Development from Non-AIDS Vaccines." *AIDS Research & Human Retroviruses* 14(Suppl 3):S197-203.

Cohen, Jon and Kai Kupferschmidt. 2014. "Ebola Vaccine Trials Raise Ethical Issues." *Science* 346(6207):289–90.

Copas, Andrew J. et al. 2015. "Designing a Stepped Wedge Trial : Three Main Designs , Carry-over Effects and Randomisation Approaches." *Trials* 16(1):1–12. Retrieved (http://www.trialsjournal.com/content/16/1/352).

Council for International Organizations of Medical Sciences (CIOMS) in collaboration with WHO. 2009. *International Ethical Guidelines for Epidemiological Studies*. Geneva. Retrieved (https://cioms.ch/wp-content/uploads/2017/01/International_Ethical_Guidelines_LR.pdf).

Council for International Organizations of Medical Sciences (CIOMS) in collaboration with WHO. 2016. *International Ethical Guidelines for Health-Related Research Involving Humans*. Geneva. Retrieved (https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf).

Cox, Nancy J. and Carolyn Buxton Bridges. 2007. "Inactivated and Live Attenuated Influenza Vaccines in Young Children." *New England Journal of Medicine* 356(7):729–31.

Dean, Natalie E., M.Elizabeth Halloran, and Ira M. Longini. 2016. "Design of Vaccine Trials during Outbreaks with and without a Delayed Vaccination Comparator." *bioRxiv* 1–18.

Diakite, Ibrahim, Eric Q. Mooring, Gustavo E. Velásquez, and Megan B. Murray. 2016. "Novel Ordered Stepped-Wedge Cluster Trial Designs for Detecting Ebola Vaccine Efficacy Using a Spatially Structured Mathematical Model." *PLoS Neglected Tropical Diseases* 10(8):1–22.

Donner, Allan and Neil Klar. 2004. "Pitfalls of and Controversies in Cluster Randomization Trials." *American Journal of Public Health* 94(3):416–22.

Doussau, Adélaïde and Christine Grady. 2016. “Deciphering Assumptions about Stepped Wedge Designs: The Case of Ebola Vaccine Research.” *Journal of Medical Ethics* 1–8.

Ebola ça Suffit Ring Vaccination Trial Consortium. 2015. “The Ring Vaccination Trial: A Novel Cluster Randomised Controlled Trial Design to Evaluate Vaccine Efficacy and Effectiveness during Outbreaks, with Special Reference to Ebola.” *British Medical Journal* 351:h3740.

Ellenberg, Susan S., Thomas R. Fleming, and David L. DeMets. 2002. *Data Monitoring Committees in Clinical Trials: A Practical Perspective*. New York: John Wiley & Sons.

Fay, Michael P., M.Elizabeth Halloran, and Dean A. Follmann. 2007. “Accounting for Variability in Sample Size Estimation with Applications to Nonadherence and Estimation of Variance and Effect Size.” *Biometrics* 63(2):465–74.

Fleming, Thomas R. 2011. “Addressing Missing Data in Clinical Trials.” *Annals of Internal Medicine* 154:113–17. Retrieved (http://www.annals.org/content/154/2/113.abstract).

Fleming, Thomas R. 2008. “Current Issues in Non-Inferiority Trials.” *Statistics in Medicine* 27(3):317–32.

Fleming, Thomas R., Katherine Odem-Davis, Mark D. Rothmann, and Yuan Li Shen. 2011. “Some Essential Considerations in the Design and Conduct of Non-Inferiority Trials.” *Clinical Trials* 8:432–39.

Fleming, Thomas R. and John H. Powers. 2012. “Biomarkers and Surrogate Endpoints in Clinical Trials.” *Statistics in Medicine* 31:2973–84.

Foege, W. H., J. D. Millar, and D. A. Henderson. 1975. “Smallpox Eradication in West and Central Africa.” *Bulletin of the World Health Organization* 52(2):209–22.

Food and Drug Administration. 2011. *Guidance for Industry: General Principles for the Development of Vaccines to Protect Against Global Infectious Diseases*.

Food and Drug Administration. 2010. *Guidance for Industry Adaptive Design Clinical Trials Guidance for Industry Adaptive Design Clinical Trials*. Silver Spring, MD.

Fraser, Christophe, Steven Riley, Roy M. Anderson, and Neil M. Ferguson. 2004. “Factors That Make an Infectious Disease Outbreak Controllable.” *Proceedings of the National Academy of Sciences* 101(16):6146–51.

Friedman, Lawrence M., Curt D. Furberg, and David L. Demets. 1998. *Fundamentals of Clinical Trials*. 3rd ed. New York: Springer Science + Business Media, LLC.

Gabriel, Erin E., Michael J. Daniels, and M.Elizabeth Halloran. 2016. “Comparing Biomarkers as Trial Level General Surrogates.” *Biometrics* 72(December):1046–54.

Gilbert, Peter B. et al. 2011. “A Sequential Phase 2b Trial Design for Evaluating Vaccine Efficacy and Immune Correlates for Multiple HIV Vaccine Regimens.” *Statistical Communications in Infectious Diseases* 3(1). Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/23181167).

Gilbert, Peter B. and Michael G. Hudgens. 2008. “Evaluating Candidate Principal Surrogate Endpoints.” *Biometrics* 64(December):1146–54.

Gilbert, Peter B., Bryan E. Shepherd, and Michael G. Hudgens. 2013. “Sensitivity Analysis of Per-Protocol Time-to-Event Treatment Efficacy in Randomized Clinical Trials.” *Journal of the American Statistical Association* 108(503):789–800.

Gill, Denis. 2004. “Ethical Principles and Operational Guidelines for Good Clinical Practice in Paediatric Research. Recommendations of the Ethics Working Group of the Confederation of European Specialists in Paediatrics (CESP).” *European Journal of Pediatrics* 163(2):53–57.

Gomes, Melba F., Vânia de la Fuente-Núñez, Abha Saxena, and Annette C. Kuesel. 2017. “Protected to Death:

Systematic Exclusion of Pregnant Women from Ebola Virus Disease Trials." *Reproductive Health* 14(S3):172. Retrieved (https://reproductive-health-journal.biomedcentral.com/articles/10.1186/s12978-017-0430-2).

Grijalva, Carlos G. et al. 2007. "Decline in Pneumonia Admissions after Routine Childhood Immunisation with Pneumococcal Conjugate Vaccine in the USA: A Time-Series Analysis." *Lancet* 369(9568):1179–86.

Gsell, Pierre Stéphane et al. 2017. "Ring Vaccination with rVSV-ZEBOV under Expanded Access in Response to an Outbreak of Ebola Virus Disease in Guinea, 2016: An Operational and Vaccine Safety Report." *The Lancet Infectious Diseases* 17(12):1276–84.

Haire, Bridget G. and Morenike O. Folayan. 2016. "Ebola 'ring' vaccine Trial Was Ethically Innovative." *American Journal of Public Health* 106(9):e1.

Halloran, M. E., I. M. Longini Jr., and C. J. Struchiner. 2010. *Design and Analysis of Vaccine Studies*. New York: Springer. Retrieved January 22, 2015 (http://www.springer.com/public+health/book/978-0-387-40313-7).

Halloran, M.Elizabeth et al. 2017. "Simulating Trials for Infectious Diseases Simulations for Designing and Interpreting Intervention Trials in Infectious Diseases Ozler Simulating Trials for Infectious Diseases." *bioRxiv* October:1–17.

Halloran, M.Elizabeth, Michael Haber, Ira M. Longini, and Claudio J. Struchiner. 1991. "Direct and Indirect Effects in Vaccine Efficacy and Effectiveness." *American Journal of Epidemiology* 133(4):323–31.

Halloran, M.Elizabeth and Ira M. Longini. 2001. "Using Validation Sets for Outcomes and Exposure to Infection in Vaccine Field Studies." *American Journal of Epidemiology* 154(5):391–98.

Halloran, M.Elizabeth, I. M. Longini Jr., Claudio J. Struchiner, Ira M. Longini, and Claudio J. Struchiner. 2010. *Design and Analysis of Vaccine Studies*. New York: Springer. Retrieved January 22, 2015 (http://www.springer.com/public+health/book/978-0-387-40313-7).

Halloran, M.Elizabeth and Claudio J. Struchiner. 1995. "Causal Inference in Infectious Diseases." *Epidemiology* 6(2):142–51.

Halloran, M.Elizabeth and Claudio J. Struchiner. 1991. "Study Designs for Dependent Happenings." *Epidemiology* 2(5):331–38.

Harling, Guy, Rui Wang, Jukka-Pekka Onnela, and Victor De Gruttola. 2017. "Leveraging Contact Network Structure in the Design of Cluster Randomized Trials." *Clinical Trials* 14(1):37–47. Retrieved (http://journals.sagepub.com/doi/10.1177/1740774516673355).

Hauck, W. W., S. Anderson, and S. M. Marcus. 1998. "Should We Adjust for Covariates in Nonlinear Regression Analyses of Randomized Trials?" *Controlled Clinical Trials* 19(3):249–56.

Hayes, R. J., N. D. Alexander, S. Bennett, and S. N. Cousens. 2000. "Design and Analysis Issues in Cluster-Randomized Trials of Interventions against Infectious Diseases." *Statistical Methods in Medical Research* 9(2):95–116. Retrieved (http://smm.sagepub.com/content/9/2/95.full.pdf).

Hayes, R. J. and S. Bennett. 1999. "Simple Sample Size Calculation for Cluster-Randomized Trials." *International Journal of Epidemiology* 28(2):319–26.

Hayes, R. J. and L. H. Moulton. 2008. *Cluster Randomised Trials*. London: Chapman & Hall/CRC.

Hayes, Richard J. and Lawrence H. Moulton. 2017. *Cluster Randomised Trials*. 2nd Editio. Boca Raton, FL: CRC Press.

Helleringer, Stephane, Patrick O. Asuming, and Jalaa Abdelwahab. 2016. "The Effect of Mass Vaccination Campaigns against Polio on the Utilization of Routine Immunization Services: A Regression Discontinuity Design." *Vaccine* 34(33):3817–22. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/27269060).

Henao-Restrepo, Ana Maria et al. 2015. “Efficacy and Effectiveness of an rVSV-Vectored Vaccine Expressing Ebola Surface Glycoprotein: Interim Results from the Guinea Ring Vaccination Cluster-Randomised Trial.” *The Lancet* 386(9996):857–66.

Henao-Restrepo, Ana Maria et al. 2017. “Efficacy and Effectiveness of an rVSV-Vectored Vaccine Preventing Ebola Virus Disease: Final Results from the Guinea Ring Vaccination, Open-Label, Cluster-Randomised Trial (Ebola Ça Suffit!).” *The Lancet* 389:505–18.

Hernán, Miguel A. and James M. Robins. 2017. “Per-Protocol Analyses of Pragmatic Trials.” *New England Journal of Medicine* 377(14):1391–98.

Higgins, Julian PT and Sally Green, eds. 2011. “16.3.2 Assessing Risk of Bias in Cluster-Randomized Trials.” in *Cochrane Handbook for Systematic Reviews of Interventions*. The Cochrane Collaboration. Retrieved (www.handbook.cochrane.org).

Hitchings, Matt D. T., Rebecca F. Grais, and Marc Lipsitch. 2016. “Using Simulation to Aid Trial Design: Ring-Vaccination Trials.” *bioRxiv*.

Horne, A. D., P. A. Lachenbruch, and K. L. Goldenthal. 2001. “Intent-to-Treat Analysis and Preventive Vaccine Efficacy.” *Vaccine* 19:319–26.

Hróbjartsson, Asbjørn et al. 2014. “Observer Bias in Randomized Clinical Trials with Time-to-Event Outcomes: Systematic Review of Trials with Both Blinded and Non-Blinded Outcome Assessors.” *International Journal of Epidemiology* 43(3):937–48.

Hudgens, Michael G., Peter B. Gilbert, and Steven G. Self. 2004. “Endpoints in Vaccine Trials.” *Statistical Methods in Medical Research* 13(2):89–114.

Hudgens, Michael G. and M.Elizabeth Halloran. 2008. “Toward Causal Inference With Interference.” *Journal of the American Statistical Association* 103(482):832–42.

Hussey, Michael A. and James P. Hughes. 2007. “Design and Analysis of Stepped Wedge Cluster Randomized Trials.” *Contemporary Clinical Trials* 28:182–91.

Institute of Medicine. 2010. *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease*. edited by C. M. Micheel and J. R. Ball. Washington, DC: National Academies Press.

International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). 2014. *E9(R1): Addendum to Statistical Principles for Clinical Trials on Choosing Appropriate Estimands and Defining Sensitivity Analyses in Clinical Trials*. Geneva, Switzerland. Retrieved (http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/E9__R1__Final_Concept_Paper_October_23_2014.pdf).

Jackson, Michael L. and Jennifer C. Nelson. 2013. “The Test-Negative Design for Estimating Influenza Vaccine Effectiveness.” *Vaccine* 31(17):2165–68. Retrieved (http://dx.doi.org/10.1016/j.vaccine.2013.02.053).

Jacobsen, Kathryn H. et al. 2016. “Lessons from the Ebola Outbreak: Action Items for Emerging Infectious Disease Preparedness and Response.” *EcoHealth* 13:200–212.

Kanapathipillai, Rupa et al. 2014. “Ebola Vaccine - An Urgent International Priority.” *New England Journal of Medicine* 371(24):2249–51.

Kennedy, Stephen B. et al. 2016. “Implementation of an Ebola Virus Disease Vaccine Clinical Trial during the Ebola Epidemic in Liberia: Design, Procedures, and Challenges.” *Clinical Trials* 13(1):49–56.

King, C. et al. 2015. “Methodological Challenges in Measuring Vaccine Effectiveness Using Population Cohorts in Low Resource Settings.” *Vaccine* 33(38):4748–55. Retrieved (http://dx.doi.org/10.1016/j.vaccine.2015.07.062).

Lachenbruch, Peter A. 1998. “Sensitivity, Specificity, and Vaccine Efficacy.” *Controlled Clinical Trials* 19(6):569–

74. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/9875836).

Lan, K. K.Gordon and David L. DeMets. 1983. “Discrete Sequential Boundaries for Clinical Trials.” *Biometrika* 70(3):659–63.

Lane, H.Clifford, Hilary D. Marston, and Anthony S. Fauci. 2016. “Conducting Clinical Trials in Outbreak Settings : Points to Consider.” 13(1):92–95.

Li, Xiaoming, Ivan S. F. Chan, and Keaven M. Anderson. 2012. “An Adaptive Design for Case-Driven Vaccine Efficacy Study When Incidence Rate Is Unknown.” *Statistics and Its Interface* 5(4):391–99.

Lipsitch, Marc et al. 2015. “Ebola and beyond.” *Science* 348(6230):46–48.

Lohr, Sharon L. 2010. *Sampling: Design and Analysis*. 2nd ed. Brooks/Cole, Cengage Learning.

Longini, I. M., M. E. Halloran, M. Haber, and R. T. Chen. 1993. “Measuring Vaccine Efficacy from Epidemics of Acute Infectious Agents.” *Statistics in Medicine* 12(3–4):249–63.

Low, Nicola et al. 2015. “A Randomized, Controlled Trial of an Aerosolized Vaccine against Measles.” *New England Journal of Medicine* 372(16):1519–29. Retrieved (http://www.nejm.org/doi/10.1056/NEJMoa1407417).

de Moraes, J.Cassio et al. 1992. “Protective Efficacy of a Serogroup B Meningococcal Vaccine in Sao Paulo, Brazil.” *The Lancet* 340(8827):1074–78.

Moulton, Lawrence H. et al. 2007. “Statistical Design of THRio: A Phased Implementation Clinic-Randomized Study of a Tuberculosis Preventive Therapy Intervention.” *Clinical Trials* 4:190–99.

Nason, Martha. 2016. “Statistics and Logistics: Design of Ebola Vaccine Trials in West Africa.” *Clinical Trials* 13(1):87–91.

National Academies of Sciences Engineering and Medicine. 2017. *Integrating Clinical Research into Epidemic Response: The Ebola Experience*. Washington, DC: National Academies Press.

National Research Council of the National Academies. 2010. *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington, DC: National Academies Press.

O’Brien, P. C. and T. R. Fleming. 1979. “A Multiple Testing Procedure for Clinical Trials.” *Biometrics* 35(3):549–56.

O’Neill, Robert T. 1988. “On Sample Sizes to Estimate the Protective Efficacy of a Vaccine.” *Statistics in Medicine* 7(12):1279–88.

Orenstein, W. a, R. H. Bernier, and a R. Hinman. 1988. “Assessing Vaccine Efficacy in the Field. Further Observations.” *Epidemiologic Reviews* 10(18):212–41.

Orenstein, WA A. et al. 1985. “Field Evaluation of Vaccine Efficacy.” *Bulletin of the World Health Organization* 63(6):1055–68. Retrieved (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1446359/pdf/11029981.pdf).

Osterholm, Michael et al. 2015. “The Ebola Vaccine Team B: A Model for Promoting the Rapid Development of Medical Countermeasures for Emerging Infectious Disease Threats.” *The Lancet Infectious Diseases* 3099(15):1–9.

Parmar, Mahesh K. B., James Carpenter, and Matthew R. Sydes. 2014. “More Multiarm Randomised Trials of Superiority Are Needed.” *The Lancet* 384(9940):283–84. Retrieved (http://dx.doi.org/10.1016/S0140-6736(14)61122-3).

Pearce, Neil. 2016. “Analysis of Matched Case-Control Studies.” *British Medical Journal* 352:i969. Retrieved (http://www.bmj.com/lookup/doi/10.1136/bmj.i969).

Perez-Heydrich, Carolina et al. 2014. “Assessing Effects of Cholera Vaccination in the Presence of Interference.” *Biometrics* 70(3):734–44.

Piantadosi, Steven. 2005. *Clinical Trials: A Methodologic Perspective: Second Edition*. Second. Hoboken, NJ: John Wiley & Sons.

Piszczek, Jolanta and Eric Partlow. 2015. “Stepped-Wedge Trial Design to Evaluate Ebola Treatments.” *The Lancet Infectious Diseases* 15(7):762–63. Retrieved (http://dx.doi.org/10.1016/S1473-3099(15)00078-X).

Pladevall, Manel et al. 2014. “Designing Multi-Center Cluster Randomized Trials: An Introductory Toolkit.” Retrieved (https://www.nihcollaboratory.org/Products/Designing CRTs-IntroductoryToolkit.pdf).

Plotkin, Stanley A. and Peter B. Gilbert. 2012. “Nomenclature for Immune Correlates of Protection After Vaccination.” *Clinical Infectious Diseases* 54(11):1615–17.

Pocock, Stuart J. 1977. “Group Sequential Methods in the Design and Analysis of Clinical Trials.” *Biometrika* 64(2):191–99.

Preziosi, Marie Pierre and M.Elizabeth Halloran. 2003. “Effects of Pertussis Vaccination on Transmission: Vaccine Efficacy for Infectiousness.” *Vaccine* 21(17–18):1853–61.

Puffer, Suezann, David Torgerson, and Judith Watson. 2003. “Evidence for Risk of Bias in Cluster Randomised Trials: Review of Recent Trials Published in Three General Medical Journals.” *BMJ (Clinical Research Ed.)* 327(7418):785–89. Retrieved (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=214092&tool=pmcentrez&rendertype=abstract).

Qin, Li, Peter B. Gilbert, Lawrence Corey, M.Juliana Mcelrath, and Steven G. Self. 2007. “A Framework for Assessing Immunological Correlates of Protection in Vaccine Trials.” *The Journal of Infectious Diseases* 196:1304–12.

Qin, Li, Peter B. Gilbert, Dean Follmann, and Dongfeng Li. 2008. “Assessing Surrogate Endpoints in Vaccine Trials with Case-Cohort Sampling and the Cox Model.” *Annals of Applied Statistics* 2(1):386–407.

Rid, Annette et al. 2014. “Placebo Use in Vaccine Trials : Recommendations of a WHO Expert Panel.” *Vaccine* 32(37):4708–12. Retrieved (http://dx.doi.org/10.1016/j.vaccine.2014.04.022).

Ridout, M. S., C. G. Demétrio, and D. Firth. 1999. “Estimating Intraclass Correlation for Binary Data.” *Biometrics* 55(1):137–48.

Rodrigues, L. C. and P. G. Smith. 1999. “Use of the Case-Control Approach in Vaccine Evaluation: Efficacy and Adverse Effects.” *Epidemiologic Reviews* 21(1):56–72.

Rothman, Kenneth J., Sander Greenland, and Timothy L. Lash. 2008. *Modern Epidemiology*. 3rd Editio. Philadelphia: Lippincott Williams & Wilkins.

Russek-Cohen, Estelle et al. 2016. “A US Food and Drug Administration Perspective on Evaluating Medical Products for Ebola.” 13(1):105–9.

Sauerwein, Robert W., Meta Roestenberg, and Vasee S. Moorthy. 2011. “Experimental Human Challenge Infections Can Accelerate Clinical Malaria Vaccine Development.” *Nature Reviews Immunology* 11(1):57–64. Retrieved (http://www.nature.com/doifinder/10.1038/nri2902).

Saxena, Abha and Melba Gomes. 2016. “Ethical Challenges to Responding to the Ebola Epidemic : The World Health Organization Experience.” *Clinical Trials* 13(1):96–100.

Schiller, John T., Xavier Castellsagué, and Suzanne M. Garland. 2012. “A Review of Clinical Trials of Human Papillomavirus Prophylactic Vaccines.” *Vaccine* 30(SUPPL.5):F123–38. Retrieved (http://dx.doi.org/10.1016/j.vaccine.2012.04.108).

Schwartz, Lauren M., M.Elizabeth Halloran, Ali Rowhani-Rahbar, Kathleen M. Neuzil, and John C. Victor. 2016. "Rotavirus Vaccine Effectiveness in Low-Income Settings: An Evaluation of the Test-Negative Design." *Vaccine*. Retrieved (http://linkinghub.elsevier.com/retrieve/pii/S0264410X16310283).

De Serres, G., D. M. Skowronski, X. W. Wu, and C. S. Ambrose. 2013. "The Test-Negative Design: Validity, Accuracy and Precision of Vaccine Efficacy Estimates Compared to the Gold Standard of Randomised Placebo-Controlled Clinical Trials." *Eurosur* 18(37):20585.

Shivayogi, Preethi. 2013. "Vulnerable Population and Methods for Their Safeguard." *Perspectives in Clinical Research* 4(1):53. Retrieved (http://www.picronline.org/text.asp?2013/4/1/53/106389).

Simon, Richard. 1989. "Optimal Two-Stage Designs for Phase II Clinical Trials." *Controlled Clinical Trials* 10(1):1–10.

Simondon, François et al. 1997. "A Randomized Double-Blind Trial Comparing a Two-Component Acellular to a Whole-Cell Pertussis Vaccine in Senegal." *Vaccine* 15(15):1606–12. Retrieved (http://www.sciencedirect.com/science/article/pii/S0264410X9700100X).

Sposto, Richard and Mark D. Krailo. 1987. "Use of Unequal Allocation in Survival Trials." *Statistics in Medicine* 6:119–25.

Sur, Dipika et al. 2009. "A Cluster-Randomized Effectiveness Trial of Vi Typhoid Vaccine in India." *New England Journal of Medicine* 361(4):355–344.

Tchetgen, E. J. T. and T. J. VanderWeele. 2012. "On Causal Inference in the Presence of Interference." *Statistical Methods in Medical Research* 21(1):55–75.

The Gambia Hepatitis Study Group. 1987. "The Gambia Hepatitis Intervention Study." *Cancer Research* 47:5782–87.

Trollfors, Birger et al. 1998. "Immunization of Children with Pertussis Toxoid Decreases Spread of Pertussis within the Family." *Pediatric Infectious Disease Journal* 17(3):196–99.

VanderWeele, Tyler J. and Eric J. Tchetgen Tchetgen. 2011. "Effect Partitioning under Interference in Two-Stage Randomized Vaccine Trials." *Statistics and Probability Letters* 81(7):861–69. Retrieved (http://dx.doi.org/10.1016/j.spl.2011.02.019).

Verani, Jennifer R. et al. 2017a. "Case-Control Vaccine Effectiveness Studies: Data Collection, Analysis and Reporting Results." *Vaccine* 35(25):3303–8. Retrieved (http://dx.doi.org/10.1016/j.vaccine.2017.04.035).

Verani, Jennifer R. et al. 2017b. "Case-Control Vaccine Effectiveness Studies: Preparation, Design, and Enrollment of Cases and Controls." *Vaccine* 35(25):3295–3302. Retrieved (http://dx.doi.org/10.1016/j.vaccine.2017.04.037).

Villar, Luis et al. 2014. "Efficacy of a Tetravalent Dengue Vaccine in Children in Latin America." *New England Journal of Medicine* 372(2):141103114505002. Retrieved (http://www.ncbi.nlm.nih.gov/pubmed/25365753).

Von, Elm E. et al. 2007. "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies." *The Lancet* 370:1453–57.

Wang, R., R. Goyal, Q. Lei, M. Essex, and V. De Gruttola. 2014. "Sample Size Considerations in the Design of Cluster Randomized Trials of Combination HIV Prevention." *Clinical Trials* 11(3):309–18.

White, Ian R., Nicholas J. Horton, James Carpenter, and Stuart J. Pocock. 2011. "Strategy for Intention to Treat Analysis in Randomised Trials with Missing Outcome Data." *British Medical Journal* 342:d40.

WHO. 2016a. *An R&D Blueprint for Action to Prevent Epidemics: Plan of Action, May 2016*. Geneva. Retrieved (http://www.who.int/csr/research-and-development/r_d_blueprint_plan_of_action.pdf?ua=1).

WHO. 2016b. “Dengue Vaccine: WHO Position Paper - July 2016.” *The Weekly Epidemiological Record* 91(30):349–64. Retrieved (http://www.who.int/immunization/policy/position_papers/dengue/en/).

WHO. 2016c. *Good Participatory Practice Guidelines for Trials of Emerging (and Re-Emerging) Pathogens That Are Likely to Cause Severe Outbreaks in the near Future and for Which Few or No Medical Countermeasures Exist*.

WHO Ethics Working Group. 2014. *Ethical Issues Related to Study Design for Trials on Therapeutics for Ebola Virus Disease. WHO Ethics Working Group Meeting 20-21 October 2014. Summary of Discussion.* Geneva.

WHO Expert Committee on Biological Standardization. 2017a. *Guidelines on Clinical Evaluation of Vaccines: Regulatory Expectations*. Retrieved (http://www.who.int/entity/biologicals/expert_committee/WHO_TRS_1004_web_Annex_9.pdf).

WHO Expert Committee on Biological Standardization. 2017b. *Human Challenge Trials for Vaccine Development: Regulatory Considerations*. Retrieved (http://www.who.int/biologicals/expert_committee/WHO_TRS_1004_web_Annex_10.pdf).

Widdowson, Marc-Alain et al. 2016. “Implementing an Ebola Vaccine Study - Sierra Leone.” *MMWR Supplements* 65(3):98–106.

World Health Organization. 2013a. *Correlates of Vaccine-Induced Protection: Methods and Implications*. WHO/IVB/13. Geneva, Switzerland: WHO Press. Retrieved (www.who.int/vaccines-documents/).

World Health Organization. 2013b. *Expert Consultation on the Use of Placebos in Vaccine Trials*. Geneva, Switzerland: WHO Press. Retrieved (http://www.who.int/ethics/publications/9789241506250/en/).

World Health Organization. 2016. *Guidance for Managing Ethical Issues in Infectious Disease Outbreaks*. Geneva, Switzerland: WHO Press.

World Medical Association. 2008. “WMA Declaration of Helsinki-Ethical Principles for Medical Research Involving Human Subjects-59th WMA General Assembly.” Retrieved (https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/).

Zou, Guang Yong, Allan Donner, and Neil Klar. 2005. “Group Sequential Methods for Cluster Randomization Trials with Binary Outcomes.” *Clinical Trials* 2(6):479–87.