

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/128334/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Hargreaves, James R., Hassan, Syreen, Schellenberg, Joanna, Hayes, Richard, Webster, Jayne and Lewis, James J. 2019. Five challenges in the design and conduct of IS trials for HIV prevention and treatment. *Journal of Acquired Immune Deficiency Syndromes* 82 , S261-S270. 10.1097/QAI.0000000000002192

Publishers page: <http://dx.doi.org/10.1097/QAI.0000000000002192>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Five challenges in the design and conduct of IS trials for HIV prevention and treatment

James R Hargreaves, PhD *

Syreen Hassan, MD *

Joanna Schellenberg, PhD *

Richard Hayes, DSc *

Jayne Webster, PhD *

James J Lewis, PhD **

Author affiliations:

*London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT

** Y Lab, the Public Services Innovation Lab for Wales, School of Social Sciences, Cardiff University

Correspondence to:

Professor James Hargreaves, PhD

London School of Hygiene & Tropical Medicine
Keppel Street
London
WC1E 7HT

E-mail: James.hargreaves@lshtm.ac.uk

Meetings at which this work was presented:

Conference on Retroviruses and Opportunistic Infections (CROI)
Seattle, USA
13-16 February 2017

Inaugural IS Network meeting, 9th International AIDS Society (IAS) Conference
Paris, France
23-26 July 2017

ViiV pre-conference workshop, 22nd International AIDS Conference
Amsterdam, the Netherlands
23-27 July 2018

Conflicts of Interest and Source of Funding:

RH receives funding from the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement, which is also part of the EDCTP2 programme supported by the European Union. Grant Ref: MR/R010161/1

The authors have no conflicts of interest or funding to disclose.

Abstract

Background: Implementation science (IS) trials in HIV treatment and prevention evaluate implementation strategies that deliver health-enhancing tools such as antiretroviral medicines or prevention technologies to those who need them, rather than evaluating the tools themselves.

Method: Opinion piece drawing on a non-systematic review of HIV prevention and treatment trials to inform an assessment of five key challenges for IS trials

Results: Randomised controlled trials (RCTs) are an appropriate design for IS but must address five challenges. IS trials must be feasible to deliver, which will require addressing challenges in maintaining multisectoral partnerships, strengthening routine data and clarifying ethical principles. IS trials should be informative, evaluating implementation strategies that are well designed and adequately described, and measuring implementation outcomes, coverage of tools and, when appropriate, epidemiological impacts. IS trials should be rigorous, striving for internally valid estimates of effect by adopting best practices, and deploying optimal non-randomised designs where randomisation is not feasible. IS trials should be relevant, considering and documenting how “real-life” is the implementation monitoring and whether research participants are representative of the target population. Finally, IS trials should be useful, deploying process evaluations to provide results that can be used in onward decision making.

Conclusion: IS trials can help ensure that efficacious tools for HIV prevention and treatment have maximum impact in the real world. These trials will be an important component of this scientific agenda if they are feasible to deliver and if their results are informative, rigorous, relevant and useful.

Key words: IS, randomised trials, HIV, HIV treatment cascade, HIV prevention cascade

Introduction

Global efforts to control the HIV epidemic have not yet maximised the population impact of prevention and treatment. Condoms prevent HIV transmission, but consistent use remains below optimal levels.¹ Regular HIV testing helps couples negotiate safer sex, and acts as a gateway to care, yet only 75% of people living with HIV (PLHIV) know their status.² Rapid scale up is needed to maximise the impact of oral pre-exposure prophylaxis (PrEP)³ and medical male circumcision.⁴ PLHIV maintained on antiretroviral therapy (ART) experience health benefits and suppression of onward transmission,³ yet globally only 47% have a controlled viral load.² Consequently, global HIV incidence remains at nearly 2 million infections a year and nearly 1 million deaths were attributable to HIV in 2017.²

The purpose of medical research should be: “to advance knowledge for the good of society; to improve the health of people worldwide; or to find better ways to treat and prevent disease.”⁵ One area of research with the potential to achieve this is IS in the field of HIV/AIDS. While definitions of IS differ, it seeks to “advance knowledge” with respect to how to organise and deliver HIV prevention and treatment services in real-world settings, such that efficacious tools (such as medicines or condoms) are used by a greater proportion of those who would benefit from them. Effective implementation strategies identified through this agenda reflect “better ways to treat and prevent disease” and have the potential to “improve the health of people worldwide”.

The range of studies that might contribute to this purpose is wide: it may, for example, include observational research to estimate the size and location of gaps in coverage of efficacious tools, and qualitative research to understand determinants of these gaps. In this

paper we will discuss one class of IS research: comparative trials of implementation strategies to deliver HIV prevention and treatment tools.

In order for the results of such trials to contribute to the purpose outlined above, we argue that scientists must address five challenges related to their design and conduct. These challenges are overlapping and sometimes obscured by current thinking. We outline the challenges and provide signposts to useful resources to navigate them. We start by providing some definitions and terminology that outline the scope of the paper and our view of the purpose of IS in HIV treatment and prevention.

Terminology

The *HIV treatment*⁶ and *prevention cascades*⁷ are critical frameworks for our view of HIV IS. These frameworks recognise that programmes impact population health through individuals' use of medicines, surgical procedures or behavioural practices that have well-estimated efficacy and safety. For the treatment cascade, this means ARTs to suppress viral load, preventing progression to AIDS and onward HIV transmission. For the prevention cascade, it might mean ART for pre- or post-exposure prophylaxis, medical male circumcision or condoms, for example. We will use the term '*tools*' to refer to these active ingredients of prevention or treatment which are efficacious when used by those who need them.

Critically, the focus of IS is not on these tools *per se*. Rather, cascades outline the steps necessary for these tools to be used by those who need them. In its simplest form the treatment cascade outlines three steps needed for ART to have population impact: first, most people living with HIV should be diagnosed; second, they should be linked to care and initiated on treatment; and, third, they should be retained in care, adhere to treatment and

thus be virally suppressed. The principles are similar for the prevention cascade. For HIV prevention tools to have population impact, first, they should be accessible; second, people at risk should be motivated to use them; and, third, those who use the tools should adhere.⁷ We will use the term *implementation strategy* to refer to any organised programmatic effort to achieve HIV prevention or treatment goals by strengthening the cascades through the deployment of, for example, health system, behavioural or structural interventions targeted at clients and/or providers.^{8,9} It is these strategies that are the focus of IS, rather than the tools.

What then is an “IS trial”? Individually randomised controlled trials (RCTs) are the gold-standard for assessing the efficacy of tools.¹⁰ RCTs involve random allocation of people to two or more “groups” that receive differing “treatments”, with outcomes then compared at a later time. We consider that an IS trial is being undertaken if the “treatment” is an implementation strategy, and if the trial measures relevant outcomes (we will discuss which outcomes may be relevant). For simplicity, we focus on randomised trials (we will briefly discuss design options when randomisation is not possible). Implementation strategies often (though not always) take the form of complex interventions organised at cluster level (a health-facility, community or some other grouping). Consequently, our discussion will focus on cluster randomised trials (CRTs).¹¹ As an example, Panel 1 highlights differences in emphasis between the HPTN052¹² and HPTN 071 (PopART) trials of a tool and an implementation strategy, respectively.^{13,14}

What then are the five challenges to the design and conduct of IS trials?

Our first challenge we label as making IS trials “feasible” to undertake. Scientists must argue for the value of investments in research, and we do not wish to limit the ambition of implementation scientists in making the case for the trials they wish to conduct.

Nevertheless, we recognise that decisions about what research to conduct are constrained by limited time and resources (especially in the context of a rapidly spreading epidemic that has most severely affected low- and middle-income countries). We will identify ways in which IS trialists face challenges in getting studies initiated and conducted in an efficient and timely manner. Our focus here is on making IS trials feasible to undertake, rather than on the feasibility of delivery of the strategies.

A second and related challenge is that of making IS trials “informative”, by which we mean challenges faced by implementation scientists in choosing what strategies to study, how to describe these (often complex) strategies to others who wish to learn from the trials they conduct, and which outcomes to measure (and, therefore, with time and resource constraints, over what time periods and scale to measure them).

A third challenge is in making the design and conduct of such trials “rigorous”, by which we mean that whatever strategies are trialled and outcomes measured, the effect sizes that are estimated should be unbiased. We will show how the real-world settings in which IS trials are often undertaken work against the desire for rigour in ways that, while not unique, are common in IS. These include standardising the application of eligibility criteria, determining the most appropriate control treatment, and situations where randomisation is not feasible or appropriate.

Fourth, we discuss the challenging decisions with respect to the design and conduct of IS trials that have implications for how “relevant” the findings are. Specifically, we will illustrate how decisions about the populations among who outcomes are studied, and the delivery and monitoring of the strategies during a trial, can influence how relevant to “real-life” the findings are perceived to be.

Fifth and finally, we discuss the challenge of making the outputs of IS trials “useful” for others. While a rigorous estimate of effect of a well described and justified implementation strategy on an outcome of interest has great value, we will show how the utility of a trial can be greatly enhanced if it is also clear how implementation was achieved in practice, how the strategy worked to achieve its outcomes (or why it did not), and what would be needed to deliver it elsewhere.

Results

Making IS trials feasible

We start by identifying three issues which can work against IS trials being undertaken at all.

First, IS trials involve partnerships between different institutions, including implementing NGOs and/or government bodies, academics, evaluators, policy makers and funders. These stakeholders have different incentives for engaging in IS trials and partnerships are fragile to maintain. In building these partnerships, one promising process in which all stakeholders can engage is in developing a theory of change using participatory approaches.^{15,16} But timing is critical: many evaluation opportunities are lost because the demands on implementers to move forward do not allow sufficient time for evaluators to engage, or because researchers’ time-intensive processes of writing protocols and seeking and receiving ethical approvals work against implementer timelines. As a trial progresses, implementers may wish to see detailed process and outcome data, but evaluators may not have the capacity to provide these and may be concerned about issues of pragmatism and fidelity to design, creating further tensions in such partnerships. There is an urgent need to develop and

strengthen platforms that bring evaluators and implementers together in ways that foster discussion and catalyse this agenda.

Second, in some settings, investigators may seek to use data from routine data systems rather than bespoke research activities to measure outcomes, but in many settings these data remain weak. We have the impression that some proponents of IS hold the view that these studies must, by definition, use only data from routine systems in tracking outcomes. We do not support this view: IS trialists must collect the data necessary to validly answer the questions they address: we therefore do not have an opposition to the collection of bespoke research data in IS trials. However, we do think that strong routine data systems would undoubtedly make it more feasible to undertake IS trials. Further, the use of routine data would enhance their value and foster greater investment in these systems. Yet current investment in routine data in many settings remains short of what is needed.

Third, IS trials pose ethical challenges. While the efficacy of the “tool(s)” delivered in such trials may be known, there should be equipoise with respect to the potential impact of new strategies on the outcomes of interest. Cluster randomised designs pose issues with regard to the level at which informed consent must be sought, the ability of participants to leave the trial, decisions about the early termination of the trial, and design and analysis considerations to ensure the scientific validity of the trial.¹⁷ Gopichandran et al. contrast the ethical requirements of IS research to clinical research, highlighting where gaps in guidance exist and the ethical issues that may arise in the planning, implementation, and post-research phases.¹⁸ In some situations, group and individual-level consent and participant tracking for follow up have been transported from the individual medicine trial to the IS trial. We do not (of course) argue for using different ethical standards in IS trials, but we caution that care is warranted if such procedures have the unintended consequence of resulting in participants

becoming unrepresentative of those who might benefit from the interventions in real-life, or undermining the conduct of trials completely. Anecdotally, we hear perceptions that ethics procedures act as a barrier to the conduct of IS trials, sometimes in ways that do not seem central to ensuring the safety of human subjects. For example, in the HPTN 071 (PopART) trial participants in Arm A were required to provide individual level research consent to receive immediate ART, despite the existence of strong individual level evidence that this was good for their own health. We would like to see the launch of a commission on ethics of IS that is wide reaching and can support university and other ethical bodies to deal with the particular challenges posed by such research.

Making IS trials informative

Implementation strategies that are evaluated in trials should be supported by a credible theory of change linking the strategy to its intended impact, justified with reference to existing knowledge, and supported by formative research. However, there are additional challenges in making design decisions about what to trial and what outcomes to measure.

One tension is in deciding how complex trialled strategies should be. Trialling individual components in isolation may not be reflective of real-life programme implementation. Each component may generate only a small change in outcomes, such that a prohibitively large trial would be needed to detect such a change if studied separately. Additionally, components may interact in non-linear ways, making it more important to study than in combination. However, where an implementation strategy contains multiple components there may be challenges in interpretation. The Botswana Combination Prevention Project (BCPP), conducted in Botswana¹⁹, and The Sustainable East Africa Research in Community Health (SEARCH) trial, conducted in Uganda and Kenya²⁰, illustrate the point. Both trials implemented complex strategies involving a range of components (see Panel 2). There were

significant overlaps, but also important differences between the interventions delivered in the intervention arms of these trials, as well as between the contexts in which they were implemented. This poses challenges for how to interpret each trial and any differences in the results between them. The SEARCH trial did not find a significantly lower HIV incidence in Kenya and Uganda in the intervention arm compared to the control, while the BCPP did in Botswana. As we will see in the next section, the availability of baseline HIV testing and rapid initiation of ART in the control group is one important factor that could explain the lack of difference in effect between the two arms of the SEARCH trial. However, the differences in the complex set of intervention components being delivered adds another layer of complexity to the interpretation of results.

Further, as interventions grow in complexity and context specificity, they become harder to describe in ways that support learning across trials. A range of tools are available to help with this. The Standards for Reporting Implementation Studies (StaRI) checklist identifies 27 items to guide the reporting of both the implementation strategy and the tool being implemented,²¹ as does the Template for Intervention Description and Replication (TIDieR).^{22,23} In developing a taxonomy for behavioural interventions, 73 discrete implementation strategies²⁴ and 93 behavioural change techniques have been identified.²⁵ The Proctor framework suggests specification of the actor, action, dose, temporality, action target, implementation outcomes, and justification for implementation strategies,²⁶ and this has been further refined using the Capability, Opportunity, Motivation, and Behaviour (COM-B) framework.^{9,27} Implementation strategies should also be grounded in behavioural theory which can help justify their likely impact and aid in interpreting findings^{28,29,24}

What outcomes should IS trials measure? Proctor proposes a set of “implementation outcomes”: acceptability, adoption, appropriateness, feasibility, fidelity, cost, penetration

and sustainability.²⁶ In some cases, it may be sufficient to identify whether different implementation strategies achieve these indicators of successful delivery. In turn, implementation trials may seek to measure the proportion of the target population who are successfully using the prevention or treatment tool. High levels of effective use of efficacious tools should translate into population impact on the epidemic, for example, mortality or HIV incidence. This conversion of efficacy (of tools) to impact (of programmes) is, we argue, the ultimate aim of most implementation strategies. But does this mean that we always need to study effects on distal outcomes in IS trials? How do we balance the need to study public health outcomes of importance with the need to avoid all trials being expensive initiatives with many years of follow up? We do not have simple answers to these complex questions. Habicht et al. make a distinction between “performance evaluations” that measure provision, utilisation, and coverage, and “impact evaluations” that measure behavioural or health outcome, suggesting the appropriate outcomes to measure depends on the policy context and decisions to be made.³⁰ Smaller trials in a range of contexts are needed to test whether novel strategies achieve implementation outcomes. However, investigators of these trials will need to be cognisant that the road to epidemiological impact cannot be assumed to be simple and this approach puts a greater emphasis on modelling projections which may carry debatable assumptions. However, the larger trials required to measure epidemiological impacts also have issues in that these huge research investments often still occur across relatively small geographies, and can take considerable time, and therefore must still confront challenges in translating their findings so that they are informative for onward decision making. There must be room in the IS agenda for trials at all points along this spectrum.

Making IS trials rigorous

RCTs are not rigorous by definition: to draw valid conclusions they must be well conducted and correctly interpreted. The design, conduct, analysis and reporting of RCTs have improved in recent years in part because of the adoption of the CONSORT statement¹⁰, which has an extension for CRTs. While IS researchers sometimes emphasise the difference between the types of trials they conduct and clinical trials, in our view CONSORT is directly relevant to the vast majority of IS trials, whether randomised or not (and similar checklists exist for non-randomised designs³¹).

Some methodological challenges will be common to many IS trials. Blinding participants and researchers to treatment arm is difficult or impossible. Consequently, an individual's eligibility for participation is often assessed after randomisation rather than before, and staff might inadvertently apply the eligibility criteria differently in the two arms. Deaton and Cartwright suggest this lack of blinding must be considered a serious threat to trials of social interventions.³² Standardised and objective measurement tools are essential to avoid biased measurement.

Another issue is in deciding what strategies to compare to the strategy of interest. A 'true' control is rarely feasible in an IS trial, since routine practice may change in the standard of care arms of such trials, as was the case with the roll-out of universal ART guidelines during the HPTN071 (PopART) trial (see Panel 1). Delaying roll out would likely be judged unethical. Many settings have other actors providing support to HIV activities that differ across geography and time, and it would be unethical to ask them to stop in order to guarantee a "true control". Given the large resources available through a trial, it is desirable, and in many cases an ethical imperative, to contribute to strengthening existing activities in the areas not receiving the intervention strategy. Some investigators may go further. In the

SEARCH trial described in Panel 2, it was decided to undertake a community wide HIV testing initiative in both intervention and control arms at the start of the trial. Initial interpretations have ascribed the lack of difference in HIV incidence between the arms to the presence of this “active control”.²⁰ How should the results of the trial be interpreted in this situation, and how can we ensure rigorous inference in such trials conducted in real-life settings? Again, this is not simple, but investigators need to carefully consider what to deliver in standard of care arms of such trials and document the components of intervention that differ between arms, as well as changes in delivery in both active and control arms of such trials over time.

Finally, in IS, randomisation may not be feasible. This paper is not the place for a discussion of reasons why this may be the case, or of alternatives to randomisation.^{33,34} Nevertheless, we stress that implementation trialists will often need to consult the tool-box of non-randomised evaluation designs.³⁵ Clear documentation of the implementation strategy and prior specification of the (non-randomised) allocation strategy allows the deployment of designs such as interrupted time series, non-randomised-controlled studies, or variations on the randomised trial such as the stepped wedge trial, or other novel approaches including adaptive designs. Although such designs have limitations, they can offer greater validity and transparency compared to observational studies.³⁵ Where this is not possible, evaluators must rely on analytic approaches to estimate effects where it is hard to avoid making multiple untestable assumptions. As for randomised trials, for these designs to produce valid results many good practices reflected in CONSORT¹⁰ or STROBE³¹, can and should be adopted.

Making IS trials relevant

In many areas of science, a continuum is recognised between *explanatory* and *pragmatic* research, which in turn relates to a distinction between *efficacy* and *effectiveness*.³⁶

Traditionally, the “efficacy” of a tool refers to how effective it is under “ideal” conditions, and it is the aim of “explanatory” research to estimate efficacy. In contrast, a tool’s “effectiveness” is thought of as its effect in “real-life” conditions, and it is the aim of “pragmatic” research to estimate such effects. We find that this distinction between efficacy and effectiveness can obscure the important difference between IS and clinical trials. In IS, we do not conceive that there are two distinct properties of a given *tool* (its efficacy in ideal conditions, and its effectiveness in the real world). Rather, IS starts from the point of view that the extent to which a tool of known efficacy has an effect in a target population depends on both the implementation strategy being deployed and the context.

One way in which IS trials seek to be relevant is by emulating “real-life” conditions. Panel 3 shows how one trial worked to ensure that the intervention was monitored as it would have been in real-life (compared to ensuring perfect fidelity with research resources), and to ensure that outcomes were measured among participants intended to benefit in real-life (compared to highly selected research participants). Current practice in trials can work against these pragmatic aims. For example, we note, anecdotally, a tendency in some settings to flexibly support interventions to maximise their chance of success in ways that can mean that they are given greater support than would be feasible in “real-life”.

Importantly, these efforts may remain undocumented as part of the implementation strategy. This can compromise the interpretation of the trial and the sustainability of the strategy when the trial comes to an end. We have also heard how processes of consent for research participants can be perceived to work against ensuring that those enrolled in research are

representative of those intended to benefit in real-life. For example, individuals from marginalized or vulnerable groups may be reticent to provide their personal details and sign “legal” consent forms. The time required to explain and obtain informed, signed, opt-in consent may be impractical in the settings of such trials.

Ensuring that trials are conducted in ways that ensure their results are relevant and interpretable is challenging, but tools are available to support trialists in designing a trial that matches how they intend to use the results.³⁷ The Pragmatic-Explanatory Continuum Indicator Summary 2 (PRECIS-2)³⁸ is a tool that encourages investigators to clarify the eligibility criteria, recruitment, setting, organisation, flexibility of delivery, and level of participant adherence to the intervention. Similarly, the CONSORT extension for pragmatic trials³⁹ requires additional information on how participants were chosen, changes that were made to the intervention in order to be implemented in the intended setting, how outcomes and length of follow up are relevant to participants and decision makers, and a discussion of the contextual elements that were essential for implementation and how they might vary in other settings. We do not suggest that all IS trials should be conducted in completely real-life conditions, but it is important for investigators and those who read their research to understand this dimension of a trial’s conduct.

Making IS trials useful

Information gained in IS trials should contribute to knowledge synthesis, help develop theory and support recommendations about what is needed to transfer effective interventions to new settings. It is commonplace to assert that while RCTs should be considered the gold standard for maximising “internal validity”, they are compromised in the extent to which they provide “external validity”.⁴⁰ We feel this distinction can obscure the more critical point for IS. All trials should be rigorous, and thus strive for internal validity. However,

understanding what can usefully be learnt from an evaluation of an implementation strategy in one setting about what is needed in new settings is not a simple case of assessing “external validity”. Rather, what is needed is a careful assessment of how and why implementation strategies are delivered and achieve their impacts (or do not) in a given setting, so that these can be considered in relation to new settings.

Consequently, undertaking an integrated process evaluation⁴¹ is essential in all IS trials²¹.

First, process evaluations should document how implementation occurred, and compare what was intended to happen with what actually happened (fidelity to design). Where interventions are flexibly delivered, documenting the process for monitoring delivery, recording which changes were made, how they were made, and by whom it was decided, allows consideration of these issues in future settings. Second, they should investigate how participants respond to the interventions in terms of their acceptability and accessibility and track behavioural pathways of change. Third, they should document salient characteristics of the context in which the implementation strategy was delivered that would influence whether the intervention works as intended. For example, what health system characteristics were necessary to support a particular intervention being delivered, or what structural factors influenced how safe sex messages could be delivered? Process evaluations are, by nature, mixed-method endeavours involving collaboration with social and behavioural scientists.⁴² STaRI includes guidance on reporting the methods and results of process evaluations. To illustrate the value of process evaluation, consider the Intervention with Microfinance for AIDS and Gender Equity (IMAGE) trial described in Panel 4.⁴³

Cartwright and Hardie seek to help policy makers improve evidence-based policy and encourage consideration of “support factors” that will help decide whether what worked “there” will work “here”.⁴⁴ Bonell argues that trials of social interventions should be seen as

“realist” as opposed to “positivist” in nature,⁴⁵ and are an ideal opportunity to study mechanisms of change and contextual factors that influence whether interventions achieve their aims in a given setting. Cartwright suggests discussing interventions at a higher level of abstraction than a detailed minutiae description of the precise components of the intervention, a way of thinking sometimes termed “middle-range theory”.⁴⁶ Finally, the Consolidated Framework for Implementation Research (CFIR) provides 39 ways to characterise five contextual domains: characteristics of the intervention; the political, social, or economic context within which implementation is occurring; the structure within which implementation will take place; characteristics of the individuals involved in the process; and the activities that allow the process of implementation to unfold, such as planning and execution.⁴⁷

Conclusion

Cutting edge science, in which randomised trials have played a critical role, has supported the identification of an array of tools that prevent or treat HIV infection if used by those who need them. We wish to see a comparable level of critical thinking given to programme implementation efforts that ensure these tools have an impact on the epidemic at large. Some of the scientific challenges in evaluating new drugs and new implementation strategies are similar and we argue that the RCT is an appropriate design in IS. Nevertheless, we have outlined challenges implementation scientists confront in undertaking trials that are feasible, informative, rigorous, relevant and useful and offer some thoughts on how to address these challenges, summarised in Table 1. IS is the critical area where progress will support achievement of the sustainable development goals, not just in relation to HIV but across all areas of global health, and we posit that IS trials will be an essential part of this scientific agenda if we can face up to these challenges.

Acknowledgements

We would like to express our gratitude to Calum Davey for his comments and constructive critique on early versions of this manuscript. We would also like to thank three reviewers and two editors (Maya Petersen and Elvin Geng) for very constructive feedback on the paper.

Tables

Panel 1: Differences in emphasis of an individually randomised clinical trial of a biomedical tool and a cluster randomised trial of an implementation strategy: the HPTN 052 and HPTN 071 (PopART) trials

The HPTN 052 trial was an individually randomised trial to establish the biological efficacy of a “tool”: ART initiated immediately regardless of CD4 count, to reduce the viral load and consequently prevent onward transmission of HIV. Its purpose was to generate scientific proof of the concept that immediate antiretroviral treatment can reduce the infectiousness of HIV-infected individuals. Newly diagnosed HIV-positive individuals in serodiscordant sexual partnerships were randomly assigned to two trial arms: immediate initiation of ART, and standard of care, which delayed the initiation of ART until CD4 count was less than 250 cells/mm³. HIV transmission rates to uninfected partners were compared between the arms. Its purpose was to generate scientific proof of the concept that immediate antiretroviral treatment can reduce the infectiousness of HIV-infected individuals. It found a 93% reduction in the sexual transmission of HIV in HIV-serodiscordant couples.¹²

By contrast, the HPTN071 (PopART) trial was a cluster randomised trial with three arms (A, B, C), randomised at community level in South Africa and Zambia, each of which reflected a different “implementation strategy”.¹³ The implementation strategy in Arm A used community HIV care providers who travelled house to house providing HIV counselling and testing, and linking HIV negative individuals to prevention services and HIV positive individuals to clinics for treatment, as a strategy to increase the use of HIV prevention and treatment tools among people. In the clinic, HIV-positive individuals were offered immediate ART regardless of their CD4 count. The strategy was identical in arm B, except that at the clinic HIV positive individuals were offered ART according to national guidelines (which changed over the course of the trial). Arm C was the standard of care arm. The HPTN071 (PopART) was designed as a CRT because the implementation strategy could not meaningfully be randomised to individual people, and there was equipoise with respect to the effects of interest on cluster-level outcomes such as community HIV incidence. The trial found a 30% and significantly lower HIV incidence in Arm B compared to Arm C, and a smaller and non-significant 7% lower HIV incidence in Arm A compared to Arm C.

Panel 2: Complex interventions and active control arms: the SEARCH and BCPP cluster randomised trials

The BCPP and SEARCH trials were cluster randomised trials of universal test and treat (UTT) strategies for HIV prevention. The BCPP trial found a 30% lower HIV incidence in the intervention compared to the control arm, while the SEARCH trial found reductions in HIV incidence in both arms, but no difference between the arms. The implementation strategies in both trials had many components, with much overlap but also some differences,⁴⁸ for example, in the SEARCH trial a multi-disease health campaign (“health fair”) was conducted, while in the BCPP trial testing was carried out for HIV only. In the SEARCH trial, the health fair and the home-based testing rounds were delivered at baseline in both intervention and control arms.

Trial	BCPP		SEARCH	
Country	Botswana		Kenya / Uganda	
HIV Prevalence	29%		4-19%	
Arm	Control	Intervention	Control	Intervention
Universal testing at baseline	-	✓ Home, mobile	✓ Multi-disease Fairs/ Home	✓ Multi-disease Fairs/ Home
Testing frequency		Baseline; ongoing; targeted	Baseline	Annual
Enhanced linkage to care		✓		✓
Rapid ART Start		✓ (from 2016)		✓
Universal Treatment regardless of CD4	✓ (from 2016)	✓ (from 2016)	✓ (from 2016)	✓
Differentiated ART Delivery				✓

Panel 3: Striving for relevance: the SAPPHIRE trial

The Sisters Antiretroviral therapy Programme for Prevention of HIV-an Integrated Response (SAPPH-IRe) trial^{49,50} was conducted to establish the impact of an enhanced package of community mobilisation and clinical service activities designed to increase female sex workers' (FSW) engagement in HIV treatment and prevention services on the proportion of all female sex workers with detectable HIV. In the intervention arm, FSW received usual care services, plus an enhanced program of community mobilisation and on-site provision of ART to increase access to and uptake of HIV treatment and care services. In the standard of care arm, the WHO standard programme for FSW⁵¹ was delivered through the national "Sisters With a Voice" programme, which involves HIV testing and counselling, referral to government services for ART, and other HIV and STI care services. The trial found improvements in viral suppression among female sex workers in both arms of the trial but no difference between the arms.

Two factors contributed to making the trial findings relevant to a "real-life" setting, as described in the table below.

Sample	The investigators sought to measure the impact of the intervention not on a convenience sample of sex workers, perhaps those most easily recruited through the clinics and likely to attend the intervention sessions. Rather, they sought to recruit a representative sample of FSW among whom there would be diverse levels of interest in, engagement with and adherence to the intervention packages. Since no sampling frame of sex workers existed, implementing this approach required the use of novel sampling strategies (respondent driven sampling), and careful analysis to report the trial aligned with CONSORT principles. ⁴⁹
Monitoring	The intervention was monitored by the implementation team using routine data and management structures as it would have been in routine practice. This meant that when these process data were analysed at the end of the trial, it showed that not all aspects of the intervention were delivered as intended. Far from reflecting a failure of the trial, this helped identify the intervention components that may not be applicable in routine practice and provided an idea of what could be more feasible for scale up.

Panel 4: The value of process evaluation: the IMAGE trial

The IMAGE intervention combined microfinance-based lending with a gender and HIV training programme aimed at empowering women in South African villages. The aims were to reduce intimate partner violence and HIV infection rates. The microfinance component provided loans to women to establish income-generating businesses and required that they meet fortnightly to manage these loans. These fortnightly meetings provided a platform for delivering the gender and HIV training programme where training sessions about gender roles, domestic violence, safe sexual practices, and HIV infection took place. Trained women would then go on to participate in community mobilisation activities to increase condom use and promote safe sexual behaviours to the young men and women of their households and communities. IMAGE reduced levels of intimate partner violence among participating women, but was not successful in changing levels of condom use or HIV infection among young people in their households or communities.⁵²

The process evaluation helped make the findings more useful, as described in the table below.⁵³

Data collection	Qualitative data: researchers' observation; focus group discussions; interviews with implementers, participants, and those who dropped out of the programme	Quantitative data: recruitment, attendance and retention
Results	<ul style="list-style-type: none"> ▶ The training programme was well attended and highly valued. ▶ Participation in community mobilisation was possible and rewarding for some women, but many were not motivated or faced difficulty in finding the time to engage in these activities. ▶ Only 10% of households in target communities directly participated in the intervention. 	
Interpretation	<ul style="list-style-type: none"> ▶ The lack of effect on condom use and HIV transmission among members of the household and community could be explained by challenges in community mobilisation, rather than by failure of the training. ▶ Interviews with providers both within and beyond the study setting revealed that the approach to training in the course of the trial was unlikely to be either sustainable <i>in situ</i> or transferrable to other microfinance organisations because of its need for high level skills and quality assurance practices that would not be widely available in many microfinance organisations. ▶ This insight made it possible to develop delivery models that would be transferrable to other settings. 	

Table 1: Summary of the challenges in the design and conduct of IS trials for HIV prevention and treatment

Challenge	Description	Suggestions for how to address the challenges
Feasibility	Partnerships for IS trials are complex and fragile; multiple stakeholders with different incentives and timelines	Develop and strengthen platforms to build partnerships (e.g. participatory approaches to co-develop Theories of Change).
	Routine data are weak; bespoke research data can be complex and expensive to collect	Investments in strengthening routine data systems are needed.
	Ethical norms have been imported from clinical and social science research and can create barriers	A commission on ethics for IS trials is needed.
Informativeness	Deciding what implementation strategies to trial and how complex they should be	Consider balance between complexity and simplicity: trialling individual components in isolation is not reflective of real-life programmes and would require large trials to detect change; multiple components pose challenges in interpretation.
	Complex implementation strategies can be difficult to describe	Use available tools to describe interventions, such as: <ul style="list-style-type: none"> ▶ The Standards for Reporting Implementation Studies (StaRI) checklist ▶ Template for Intervention Description and Replication (TIDieR) ▶ The Proctor framework: specification of the actor, action, dose, temporality, action target, implementation outcomes, and justification for implementation strategies.
	Which outcomes to measure? Not all trials can be very large and study both proximate and distal outcomes	<ul style="list-style-type: none"> ▶ Consider the policy context and what decisions are to be informed or strengthened by evidence ▶ Consider balance: smaller trials in a range of contexts are needed to measure the effect of strategies on implementation outcomes but requires strong assumptions for epidemiological impact; larger trials are needed to measure epidemiological impacts but occur across small geographies and can take considerable time.
Rigor	Are the rules of RCTs relevant to IS trials?	Most of the time, the CONSORT can be used to guide IS trials.

	Lack of blinding of participants; eligibility assessed after randomisation; risk of applying the eligibility criteria differently in the two arms	Use of standardised and objective measurement tools to avoid biased measurement.
	Lack of a “true control”	<ul style="list-style-type: none"> ▶ Careful consideration of what to deliver in the standard of care arm ▶ Document the components that differ between arms, and the changes in delivery in both active and control arms over time.
	Randomisation may not be feasible in some settings	<ul style="list-style-type: none"> ▶ Use quasi-experimental and alternative trial designs ▶ Clear documentation of the implementation strategy and prior specification of the non-randomised allocation
Relevance	The distinction between efficacy and effectiveness is not clear in IS	<ul style="list-style-type: none"> ▶ The extent to which a tool of known efficacy has an effect in a target population depends on both the implementation strategy and the context ▶ Focus on the effectiveness of the implementation strategy rather than the efficacy of the tool
	Deciding how pragmatic the trial should be is not simple	Where possible, recruit populations intended to benefit in real-life, monitor the intervention as it would be in routine practice, and use tools (e.g. PRECIS-2) to match design and conduct to the intended use of results.
Usefulness	What happens in practice may differ from the intended implementation	Integrate a process evaluation to document how implementation occurred, how participants respond to the interventions, and the salient characteristics of the context in which the implementation strategy was delivered
	Clients may respond in a range of ways to interventions	
	How to learn from an evaluation of an implementation strategy in one setting about what is needed in new settings?	

References

1. Steiner MJ, Cates W, Warner L. The real problem with male condoms is nonuse. *Sex Transm Dis*. 1999;26(8):459-462.
2. UNAIDS. *AIDS Statistics—2018 Fact Sheet*. Geneva: UNAIDS; 2018.
3. World Health Organisation. *Guidance on Couples HIV Testing and Counselling Including Antiretroviral Therapy for Treatment and Prevention in Serodiscordant Couples: Recommendations for a Public Health Approach*. Geneva: WHO; 2012.
4. Weiss HA, Quigley MA, Hayes RJ. Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis. *Aids*. 2000;14(15):2361-2370.
5. What is the purpose of medical research? *Lancet*. 2013;381(9864):347.
doi:10.1016/S0140-6736(13)60149-X
6. Gardner EM, McLees MP, Steiner JF, Del Rio C, Burman WJ. The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. *Clin Infect Dis*. 2011;52(6):793-800. doi:10.1093/cid/ciq243
7. Hargreaves JR, Delany-Moretlwe S, Hallett TB, et al. The HIV prevention cascade: integrating theories of epidemiological, behavioural, and social science into programme design and monitoring. *Lancet HIV*. 2016;3(7):e318-e322.
doi:10.1016/S2352-3018(16)30063-7
8. Krishnaratne S, Hensen B, Cordes J, Enstone J, Hargreaves JR. Interventions to strengthen the HIV prevention cascade: a systematic review of reviews. *Lancet HIV*. 2016;3(7):e307-e317. doi:10.1016/S2352-3018(16)30038-8

9. Hickey MD, Odeny TA, Petersen M, et al. Specification of implementation interventions to address the cascade of HIV care and treatment in resource-limited settings: A systematic review. *Implement Sci.* 2017;12(1):1-15. doi:10.1186/s13012-017-0630-8
10. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Med.* 2010;8(1):18.
11. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Heal Care.* 2003;12. doi:10.1136/qhc.12.1.47
12. Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 Infection with Early Antiretroviral Therapy. *N Engl J Med.* 2011;365(6):493-505. doi:10.1056/NEJMoa1105243
13. Hayes R, Ayles H, Beyers N, et al. HPTN 071 (PopART): rationale and design of a cluster-randomised trial of the population impact of an HIV combination prevention intervention including universal testing and treatment - a study protocol for a cluster randomised trial. *Trials.* 2014;15:57. doi:10.1186/1745-6215-15-57
14. Hayes RJ, Donnell D, Floyd S, et al. Effect of Universal Testing and Treatment on HIV Incidence — HPTN 071 (PopART). *N Engl J Med.* 2019;381(3):207-218. doi:10.1056/NEJMoa1814556
15. Breuer E, De Silva MJ, Fekadu A, et al. Using workshops to develop theories of change in five low and middle income countries: lessons from the programme for improving mental health care (PRIME). *Int J Ment Health Syst.* 2014;8(1):15.
16. Peters DH, Bhuiya A, Ghaffar A. Engaging stakeholders in implementation research:

- lessons from the Future Health Systems Research Programme experience. 2017.
17. Hutton JL. Are distinctive ethical principles required for cluster randomised controlled trials? *Stat Med.* 2001;20. doi:3.0.CO;2-D
 18. Gopichandran V, Luyckx VA, Biller-Andorno N, et al. Developing the ethics of implementation research in health. *Implement Sci.* 2016;11(1):161. doi:10.1186/s13012-016-0527-y
 19. Makhema J, Wirth KE, Pretorius Holme M, et al. Universal Testing, Expanded Treatment, and Incidence of HIV Infection in Botswana. *N Engl J Med.* 2019;381(3):230-242. doi:10.1056/NEJMoa1812281
 20. Havlir D V., Balzer LB, Charlebois ED, et al. HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa. *N Engl J Med.* 2019;381(3):219-229. doi:10.1056/NEJMoa1809866
 21. Pinnock H, Barwick M, Carpenter CR, et al. Standards for Reporting Implementation Studies (StaRI) Statement. *BMJ.* 2017;356. doi:10.1136/bmj.i6795
 22. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. doi:10.1136/bmj.g1687
 23. Cotterill S, Knowles S, Martindale AM, et al. Getting messier with TIDieR: Embracing context and complexity in intervention reporting. *BMC Med Res Methodol.* 2018;18(1):12. doi:10.1186/s12874-017-0461-y
 24. Powell BJ, Waltz TJ, Chinman MJ, et al. A refined compilation of implementation strategies: Results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement Sci.* 2015;10(1):1-14. doi:10.1186/s13012-015-0209-1
 25. Michie S, Carey RN, Johnston M, et al. From Theory-Inspired to Theory-Based

- Interventions: A Protocol for Developing and Testing a Methodology for Linking Behaviour Change Techniques to Theoretical Mechanisms of Action. *Ann Behav Med.* 2016;1-12. doi:10.1007/s12160-016-9816-6
26. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. *Implement Sci.* 2013;8(1):139. doi:10.1186/1748-5908-8-139
 27. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci.* 2011;6. doi:10.1186/1748-5908-6-1
 28. Birken SA, Powell BJ, Shea CM, et al. Criteria for selecting implementation science theories and frameworks: Results from an international survey. *Implement Sci.* 2017;12(1):1-9. doi:10.1186/s13012-017-0656-y
 29. Nilsen P. Making sense of implementation theories, models and frameworks. *Implement Sci.* 2015;10:53. doi:10.1186/s13012-015-0242-0
 30. Habicht JP, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *Int J Epidemiol.* 1999;28(1):10-18. doi:10.1093/ije/28.1.10
 31. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol.* 2008;61(4):344-349. doi:10.1016/j.jclinepi.2007.11.008
 32. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med.* 2018;210:2-21. doi:10.1016/j.socscimed.2017.12.005
 33. Bonell CP, Hargreaves J, Cousens S, et al. Alternatives to randomisation in the

- evaluation of public health interventions: Design challenges and solutions. *J Epidemiol Community Health*. 2011;65(7). doi:10.1136/jech.2008.082602
34. Cousens S, Hargreaves J, Bonell C, et al. Alternatives to randomisation in the evaluation of public-health interventions: Statistical analysis and causal inference. *J Epidemiol Community Health*. 2011;65(7). doi:10.1136/jech.2008.082610
 35. Davey C, Boulay M, Hargreaves JR. Strengthening Nonrandomized Studies of Health Communication Strategies for HIV Prevention. *JAIDS J Acquir Immune Defic Syndr*. 2014;66:S271-S277. doi:10.1097/QAI.0000000000000232
 36. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol*. 2009;62(5):464-475.
 37. Zwarenstein M, Treweek S, Loudon K. PRECIS-2 helps researchers design more applicable RCTs while CONSORT Extension for Pragmatic Trials helps knowledge users decide whether to apply them. *J Clin Epidemiol*. 2017;84(Supplement C):27-29. doi:https://doi.org/10.1016/j.jclinepi.2016.10.010
 38. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *bmj*. 2015;350:h2147.
 39. Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ*. 2008;337:a2390. doi:10.1136/BMJ.A2390
 40. Rothwell PM. External validity of randomised controlled trials:“to whom do the results of this trial apply?” *Lancet*. 2005;365(9453):82-93.
 41. Moore G, Audrey S, Barker M, Bonell C, Hardeman W, Moore L, O’Cathain A, Tinati

- T, Wight D BJ. *Process Evaluation of Complex Interventions: UK Medical Research Council (MRC) Guidance.*; 2014.
42. Palinkas L a, Horwitz SM, Chamberlain P, Hurlburt MS, Landsverk J. Mixed-methods designs in mental health services research: a review. *Psychiatr Serv.* 2011;62(3):255-263. doi:10.1176/appi.ps.62.3.255
 43. Intervention T, Study GE, Study I, Africa S. Study Protocol The Intervention with Microfinance for AIDS and Gender Equity Study (IMAGE Study) : An integrated community randomised trial of a structural intervention to prevent HIV and gender-based violence in South Africa. *Health Policy (New York).* 2005;1:1-68.
 44. Cartwright N, Hardie J. *Evidence-Based Policy: A Practical Guide to Doing It Better.* Oxford University Press; 2012.
 45. Bonell C, Fletcher A, Morton M, Lorenc T, Moore L. Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Soc Sci Med.* 2012;75(12):2299-2306. doi:10.1016/j.socscimed.2012.08.032
 46. Davey C, Hargreaves J, Hassan S, et al. Designing evaluations to provide evidence to inform action in new settings. 2018.
 47. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci.* 2009;4. doi:10.1186/1748-5908-4-50
 48. Perriat D, Balzer L, Hayes R, et al. Comparative assessment of five trials of universal HIV testing and treatment in sub-Saharan Africa. *J Int AIDS Soc.* 2018;21(1). doi:10.1002/jia2.25048

49. Hargreaves JR, Fearon E, Davey C, Phillips A, Cambiano V, Cowan FM. Statistical design and analysis plan for an impact evaluation of an HIV treatment and prevention intervention for female sex workers in Zimbabwe: a study protocol for a cluster randomised controlled trial. *Trials*. 2016;17:6. doi:10.1186/s13063-015-1095-1
50. Cowan FM, Davey C, Fearon E, et al. Targeted combination prevention to support female sex workers in Zimbabwe accessing and adhering to antiretrovirals for treatment and prevention of HIV (SAPPH-IRe): a cluster-randomised trial. *Lancet HIV*. 2018;5(8):e417-e426. doi:10.1016/S2352-3018(18)30111-5
51. WHO. *Preventing HIV in Sex Work Settings in Sub-Saharan Africa*. World Health Organization; 2011.
52. Pronyk PM, Hargreaves JR, Kim JC, et al. Effect of a structural intervention for the prevention of intimate-partner violence and HIV in rural South Africa: a cluster randomised trial. *Lancet*. 2006;368(9551). doi:10.1016/S0140-6736(06)69744-4
53. Hargreaves J, Hatcher A, Strange V, et al. Process evaluation of the Intervention with Microfinance for AIDS and Gender Equity (IMAGE) in rural South Africa. *Health Educ Res*. 2010;25(1). doi:10.1093/her/cyp054