DESIGN, ANALYSIS, AND REPORTING OF PILOT STUDIES IN HIV

DESIGN, ANALYSIS, AND REPORTING OF PILOT STUDIES IN HIV

BY

HUSSEIN ALI EL-KHECHEN, B.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF HEALTH RESEARCH

METHODS, EVIDENCE AND IMPACT

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

TITLE: Design, Analysis, and Reporting of Pilot Studies in HIV

AUTHOR: Hussein Ali El-Khechen

B.Sc. (Human Biology and Psychology),

McMaster University, Hamilton, Canada

SUPERVISOR: Dr. Lawrence Mbuagbaw

NUMBER OF PAGES: xi, 48

Lay Abstract

Pilot studies are important in evaluating whether planned larger studies can be conducted. They are particularly useful in the field of HIV where participants may be hard to identify and recruit. However, there are few instructions on how pilot studies in HIV should be designed. We searched the literature to see the current state of HIV pilot studies, including how they are designed, and their findings reported. We found that pilot studies are becoming more popular in the HIV field. However, there were gaps in how these studies are designed and reported. Studies were often mislabeled as pilots when they were not, the pilot study criteria were applied inconsistently and the outcomes that were evaluated were often poorly defined and their information poorly presented. Pilot studies in HIV can be reported better.

Abstract

Pilot studies, a subset of feasibility studies, are essential in determining the feasibility of a larger study. This is especially true when targeting populations that are difficult to recruit, such as people with HIV. Designing high quality pilot studies can help limit waste by informing researchers how to proceed.

We conducted a meta-epidemiological review of pilot studies in the HIV literature published until November 25, 2020 using Medline, Embase and Cochrane Controlled Register of Trials (CENTRAL). We extracted bibliometric information, including the region and income of the country where the study was conducted, study design, using the pilot label, source of funding, nature of intervention, whether feasibility was the primary objective, progression criteria, protocol registration and sample size estimation. We used descriptive analysis to evaluate how pilot studies are designed and conducted, the outcomes assessed and how are they defined.

Our search retrieved 10,597 studies, of which 248 were included in our final review. The number of pilot studies has increased with time, with 25, 55, and 44 HIV studies published in 2018, 2019, and 2020, respectively. We found that 128 studies (70.39%) used the pilot or feasibility labels in their title, however 20.31% used these titles interchangeably. 5 studies in this review included progression criteria, all of which were published in 2020. Sample size estimation was only found in 59 studies (23.9%).

Pilot studies in the HIV literature are mislabeled. Sample size estimations are seldom included, and progression criteria are used. Formal guidance on the design and reporting of pilot studies in the HIV literature is necessary.

Dedication

To my parents- Thank you for nursing me with love and care and teaching me to make the best out of any situation. After all, any struggle is a chance to grow and get stronger. My grandparents were poor farmers, and your dreams and education were taken from you due to war. This degree is not only for myself, but for you and my grandparents (may God rest their souls). You instilled in me the desire to get better and the importance of higher education. You lived through war, left your families, sacrificed your health, and gave everything to myself and my siblings so we can have the opportunity you did not have. For that, I am eternally grateful, and I can never repay you, even in 10 lifetimes. Everything I do is to make you proud.

To Priya, Usman, and Sukrit- Thank you for always being in my corner and checking on me, especially when I would be holed up and in need of a pick me up. From working on the innovation sprint competition to chugging coffee, hanging out in the HEI room jamming to ABBA, and laughing uncontrollably while trying to finish projects, the memories we made were the best part of the last two years. You constantly pushed me to be the best version of myself I can be and believed in me when I did not. For that, I am forever grateful. And Priya, please remind me to never eat multi-grain toast again when a deadline is fast approaching. I am leaving HRM with another family, because our friendship is the type to last a lifetime. I am excited to see each of us grow and to enjoy the good times ahead with you!

To my customers- Thank you for your kindness, encouragement, and checking in on me, especially when I would be typing away while at work. The interest in my work that each of you has shown me has been an immense source of motivation. From Bay Street bankers, healthcare workers in the surrounding hospitals and the panhandlers I got to know, each one of you took a genuine interest in my studies and showed me that there is goodness everywhere. I could not have done it without you.

Acknowledgement

I would like to express my deepest gratitude to my supervisor, Dr. Lawrence Mbuagbaw for his guidance and support. This thesis would not have been possible were it not for his inspiring optimism and contagious calmness, even when facing the greatest of challenges. You took a chance of me and I am forever in your debt. My only hope is that I can live up to your expectations and make you proud.

Contents

List of Figures

List of Tables

Abbreviations

NIHR	National Institute for Health Research
RfPB	Research for Patient Benefit
HIV	Human Immunodeficiency Virus
CENTRAL	Cochrane Central Register of Controlled Trials
WHO	World Health Organization
RCT	Randomized Controlled Trial
ART	Antiretroviral Therapy
ACB	African, Caribbean, and Black
MSM	Men who have Sex with Men
PWID	People Who Inject Drugs
CSW	Commercial Sex Workers
STROBE	STrengthening the Reporting of Observational studies in Epidemiology
NGT	Nominal Group Technique

Introduction

Wastefulness in medical research is a major concern for researchers and funders and has been estimated to be at 85% of research investment (Glasziou, 2014). There are several contributors to this waste. These include researchers not asking relevant questions, study results being inaccurately reported, and the inappropriate use of study designs (Chalmers & Glasziou, 2009). Recent work has demonstrated that pilot studies are very effective in reducing waste and their use should be more widespread (Morgan, Hejdenberg, Hinrichs-Krapels, & Armstrong, 2018). It was found that by employing pilot studies, the UK's National Institute for Health Research's (NIHR) Research for Patient Benefit (RfPB) programme saved approximately £20m, as otherwise non-feasible studies would have been conducted (Morgan et al., 2018). Pilot studies are especially useful in fields where participants are difficult to recruit and retain. However, despite the recognized value of pilot studies, there is still considerable confusion surrounding what constitutes a pilot study, how they should be designed, and how researchers decide whether they should proceed with the full study.

A pilot study is often described as a scaled down version of a larger study, with feasibility i.e. the assessment of the ability to conduct the full-scale study as the primary goal (Eldridge, Lancaster, et al., 2016). Findings from these studies are often used to inform an investigator's decision on how to proceed with the study. However, this should not be confused with a feasibility study (Eldridge, Lancaster, et al., 2016). According to Eldridge et al., "a feasibility study asks whether

something can be done, should we proceed with it, and if so, how,"(Eldridge, Lancaster, et al., 2016). Pilot studies build on this by specifying the study design by which the feasibility question will be assessed (Eldridge, Lancaster, et al., 2016). The terms "pilot" and "feasibility" continue to be used interchangeably; however, pilot studies are defined as being a subset of feasibility studies (Eldridge, Lancaster, et al., 2016). Researchers often use this label inappropriately for studies with small sample sizes (Bugge et al., 2013). In fact, there is evidence suggesting that only 56% of pilot and feasibility studies discuss feasibility outcomes and methodological issues (Milensu Shanyinde, Ruth M. Pickering, & Mark Weatherall, 2011). It is important for pilot studies to have a feasibility outcome as their primary outcome.

As the primary focus of pilot studies, feasibility outcomes can be tailored to the needs of a study. Researchers can evaluate any aspect of a study, resulting in many possibilities for feasibility outcomes. Recognizing the challenge of collecting and defining all the outcomes, there are currently efforts to group feasibility outcomes into categories instead. Some researchers have suggested categories based on the most common outcomes, such as acceptance and implementation (Bowen et al., 2009). However, other researchers group these outcomes by the 4 reasons to conduct a pilot study (Thabane et al., 2010). These include 1) assessing the processes involved in the study (ex. recruitment rates), 2) evaluation of resources required for the study (ex. retention rates, refusal rates, and failure/success rates), 3) management of potential human and data management problems (ex. are study centers able to adhere to study protocol), and 4) assessment

of intervention safety, dose response, and variance of effect (ex. what is a safe dose) (Thabane et al., 2010). Given the broad scope of feasibility outcomes, it is evident that researchers must clearly specify and define their outcomes.

Researchers rely on feasibility data to determine whether they should proceed with a larger trial. Currently, there is no information on what influences this decision. Instead, Thabane et al. and Bugge et al. outline three paths forward after the completion of a pilot study (Bugge et al., 2013; Thabane et al., 2010). Firstly, authors may determine that a larger trial is not feasible and should be terminated after the pilot stage (Bugge et al., 2013; Thabane et al., 2010). Secondly, both agree that investigators may also adapt their studies using the findings of the pilot (Bugge et al., 2013; Thabane et al., 2010). Alterations based on the results of the pilot study could be applied to the intervention, the clinical context, or a combination of several factors (Bugge et al., 2013; Thabane et al., 2010). Thirdly, trialists may elect to continue without modifying the trial, but monitor it closely (Bugge et al., 2013; Thabane et al., 2010). While these suggestions are helpful in guiding authors upon the completion of their pilot study, they do not inform us on what factors impact these decisions.

Pilot Studies in HIV

Given the numerous challenges of recruiting and retaining participants in HIV research important in this field (J. K. Anastasi, B. Capili, G. H. Kim, & A. Chung, 2005; Cook, Mack, & Cottler, 2018; De La Rosa, Babino, Rosario, Martinez, & Aijaz, 2012; el-Sadr & Capps, 1992; Fortune, Wright, Juzang, & Bull,

2010; Hoffman et al., 2019; Menezes et al., 2011; Nalubega & Evans, 2015; Silvestre et al., 2006; Yehia et al., 2015), it is natural for pilot studies to be employed in HIV research. Pilot studies help researchers understand how they can better adjust their procedures and reduce waste, especially when working with populations that are difficult to study (Morgan et al., 2018). People with HIV also often belong to other key populations which may face additional social stigma (e.g. men having sex with men (MSM), commercial sex workers, or injection drug users), making recruitment more difficult (J. K. Anastasi et al., 2005; De La Rosa et al., 2012; el-Sadr & Capps, 1992; Fortune et al., 2010; Gemmill, Williams, Cooke, & Grant, 2012; Silvestre et al., 2006). Studies have found that the general population and key subpopulations (women, people who inject drugs (PWID), and African, Caribbean and Black (ACB) peoples) have dropout rates ranging from 30%-50% (Joyce K. Anastasi, Bernadette Capili, Gee H. Kim, & Ann Chung, 2005; Asiimwe, Kanyesigye, Bwana, Okello, & Muyindike, 2016; Batista et al., 2016; Loutfy et al., 2014). However, it is important to note that participant mental health, sex, age, education level and race have also been found to be highly predictive of participant retention (Bulsara, Wainberg, & Newton-John, 2018). Finally, HIV studies are often multicentered, and as a result, there may be inconsistencies in adherence to protocol and recruitment and retention of participants between centers (Irving & Curley, 2008). As a result, pilot studies may help improve the feasibility of larger studies.

As we see above, studying HIV patients is difficult given the challenges in recruitment and retention, as well as maintaining fidelity to the study protocol and consistency between centers. Therefore, it is unsurprising that pilot studies are common in this area (J. K. Anastasi et al., 2005; Cook et al., 2018; De La Rosa et al., 2012; el-Sadr & Capps, 1992; Fortune et al., 2010; Hoffman et al., 2019; Irving & Curley, 2008; Menezes et al., 2011; Minisman et al., 2012; Nalubega & Evans, 2015; Silvestre et al., 2006; Yehia et al., 2015). To provide guidance, it is important to have an accurate assessment of pilot studies in the HIV literature. And so, we set out to fill this gap in the literature. The primary goal of this study was to answer the question "what is the current state of pilot studies in the HIV literature?" This was defined as how 1) pilot studies are labeled in their titles and abstracts, 2) how are studies designed and conducted, 3) whether they abide by the pilot study criteria, 4) what are the outcomes assessed and how are they defined, and 5) how feasibility outcome data is reported. To obtain this information, we conducted a thorough methodological study of pilot studies in the HIV literature.

Methods

We conducted a methodological study of pilot studies in the HIV literature as per the guidelines reported by Murad and Wang for reporting meta-epidemiological research (Murad & Wang, 2017).

Criteria for Inclusion

All pilot studies of interventions conducted exclusively in people with HIV were included in this review. Randomized and non-randomized studies labeled as pilot, feasibility or exploratory studies were included. Only studies employing quantitative or mixed methods were included. Studies must have been of a clinical interventional nature and assessing at least one feasibility outcome (Thabane et al., 2010). Outcomes were deemed to be feasibility outcomes if they fit into a category outlined by Thabane et al. (Thabane et al., 2010). These categories include 1) assessing the processes involved in the study, 2) evaluation of resources required for the study, 3) management of potential human and data management problems, and 4) assessment of intervention safety, dose response, and variance of effect.

Search Method for Identifying Pilot Studies

We conducted an exhaustive search of the following databases: Medline, Embase and the Cochrane Central Register of Controlled Trials (CENTRAL). These databases were searched from inception to November 25, 2020. An English language restriction was applied. Our Medline search strategy was developed in collaboration with a librarian at the library services of the McMaster Health Sciences Central Library. The search strategy was later adapted to the other databases with the analogous subject headings and keywords used for each database. The key concepts included in the search were “pilot,” “feasibility,” “proof-of-concept,” “exploratory,” “preliminary”, and “HIV.”

Pilot Study Selection, Data Collection and Data Management

We compiled the references and removed duplicate citations using the Endnote X9 reference manager software (Hupe, 2019). References deemed by Endnote X9 to have 50% similarity or more were evaluated individually. We screened the remaining references first by their title and abstracts and then by examining their electronic full texts collected from the McMaster University Health Sciences Library. Both screening steps were done in duplicate by two independent reviewers using the Covidence online screening service provided by McMaster University (Kellermeyer, Harnke, & Knight, 2018). The reviewers attempted to resolve discrepancies once the screening was completed. In cases that this did not work, a third reviewer helped resolve the discrepancy.

Data from references deemed to fulfill the inclusion criteria were extracted using a piloted data-extraction form on RedCap (Harris et al., 2009). Basic bibliometric information extracted from the studies included the following: the first author's last name, year of publication, journal of publication, and country of study (both region and income level). Region was determined using the regional groupings definitions provided by the World Health Organization (WHO) and income level was determined as per the World Bank Criteria (Bank, 2018; WHO, 2020). Other information collected included: study title, objectives, source of funding, trial design, the population included (inclusion of key populations), intervention type, whether feasibility was a primary outcome, whether the terms pilot and/or feasibility were used in the title, the type of feasibility outcome(s)

assessed and how they were defined, analysis of feasibility outcomes conducted (qualitative, quantitative, or mixed methods), the progression criteria specified, the study sample size and how it was reported, and the authors' decision on whether to proceed with the larger study. The key populations were identified using categories established in the HIV literature and by the Joint United Nations Programme on HIV/AIDS (UNAIDS) and WHO (Bekker, Johnson, Wallace, & Hosek, 2015; Djomand, Quaye, & Sullivan, 2014; Macdonald, Verster, & Baggaley, 2017; Rao et al., 2017). The key populations for which data was collected on included 1) injection drug users, 2) MSM, 3) incarcerated populations, 4) commercial sex workers (CSW), 5) pregnant women, 6) children, 7) youth, 8) indigenous people, 9) ACB people, 10) women, 11) transgender people. We contacted authors via email to clear up ambiguity or to collect missing data.

Data Synthesis and Interpretation

We conducted a descriptive analysis and reported counts and percentages for categorical variables and median (minimum, maximum) for continuous variables.

Ethics

This study used only secondary publicly available data and therefore ethics review was not required.

Results

Results of search

Our search returned 10,597 articles for title and abstract review. Of these, 386 were retrieved for full-text review as they met our inclusion criteria or required further evaluation to assess eligibility where there was ambiguity. During the full-text screening stage, an additional 288 studies were excluded. 125 were found to not meet our inclusion criteria, with 81 studies not evaluating a feasibility outcome. The remaining 44 studies either were not interventional or clinical in nature, included non-HIV patients, or were entirely qualitative studies. Figure 1 is a flow diagram of the screening process. The primary reason for reference exclusion was due to references being published or conference abstracts, study protocols or due to an inability to retrieve the full text (k=163). 2 references were excluded for being duplicate publications and one for being written in the Danish language. Our final methodological study included 248 publications.

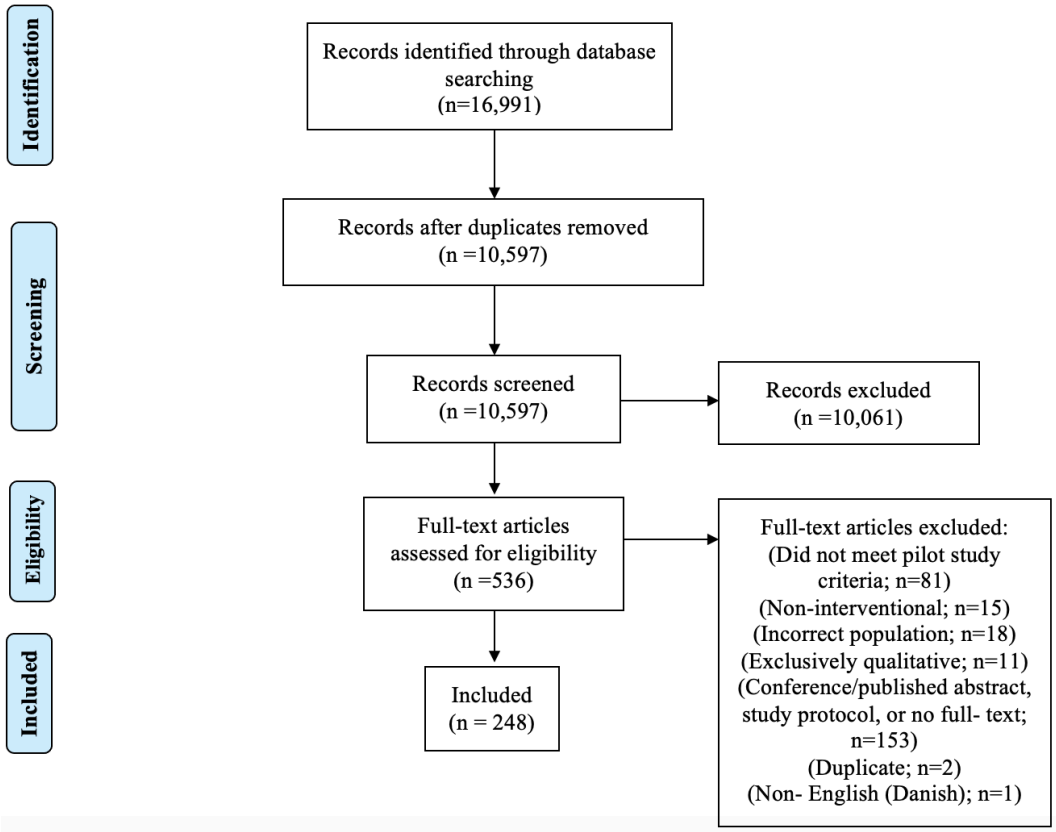

Figure 1: The number of studies identified, excluded and included in the analysis. Our searched returned 16,991 items to be screened. After deduplication, 10,597 items underwent title and abstract screening. 536 of these were either suitable for inclusion or required additional scrutiny. 248 manuscripts were included.

Study Characteristics

The one hundred and seventy-nine studies included in our review were published between 1998-2020. However, the study by Newell et al. in 1998 is an outlier, with the next study being published in 2007. Since then, there has been a steady increase in the publication of clinical interventional pilot studies in the HIV literature, with 2- 22 publications each year (Figure 2). However, 2019 and 2020 showed a dramatic increase, with 55 and 44 pilot studies published, respectively.

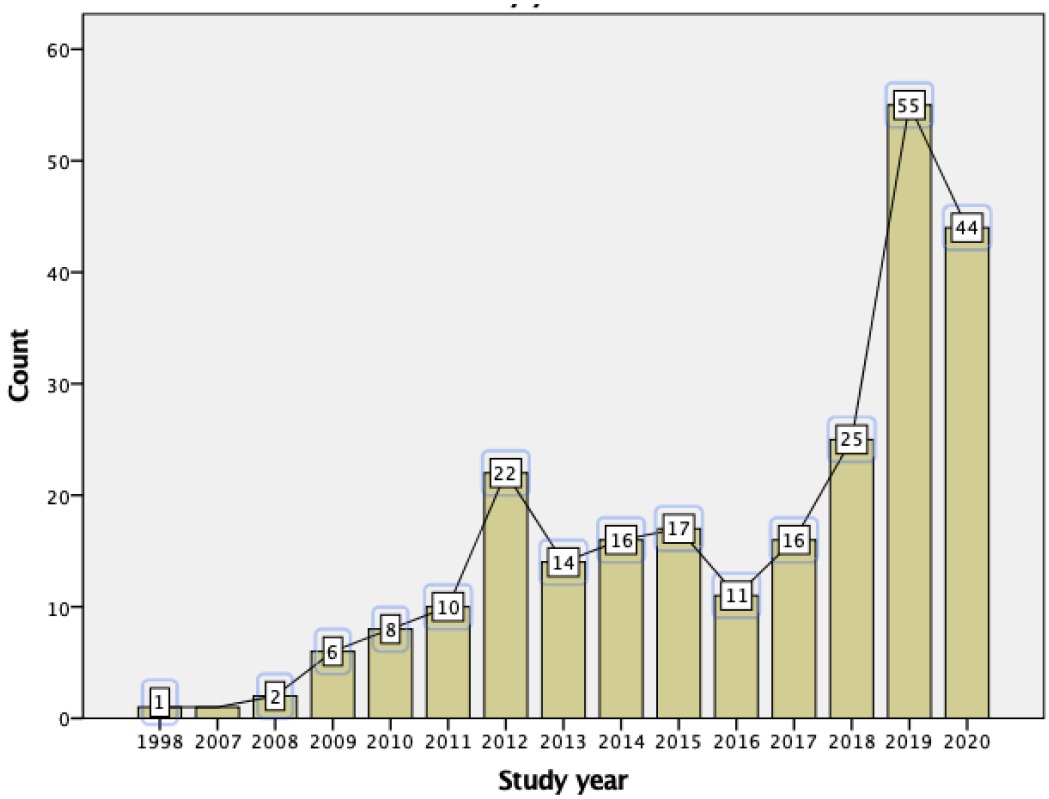

Figure 2: Publication of pilot studies in the HIV literature over the years. With the exception of a gap between 1998 and 2007, there has been a steady increase in the number if pilot studies in the HIV literature. The 2 most recent years, 2019 and 2020, have seen a significant increase (k=55 and 44, respectively).

The total sample size of the studies included in this review is 40,534 people with HIV, with a median sample size of 40 patients per study (minimum n=3, maximum n=8,794). Authors often recruited diverse patient populations, with the proportion of HIV patients belonging to a vulnerable population group in these studies has also steadily increasing with time (Figure 3). Since 2018, close to half of studies published have contained vulnerable populations, with these proportions being 60% in 2018, 47% in 2019, and 52.2% in 2020. A sample size estimation was provided in 59 studies (23.8%). Authors provided a wide range of justifications for their sample size estimations. Sample size estimations were primarily based on the

intervention's effect, which was retrieved from the literature (k=20, 33.9%), with the remaining basing their estimations on other similar studies (k=6, 10.2%), a fixed proportion of the larger study's sample size (k=2, 3.4%), recommendations in the literature (k=6 ,2.4%), or other reasoning was provided (k=5, 8.5%). However, the authors of 17 studies (28.8%) did not provide justification for their prespecified sample size estimations.

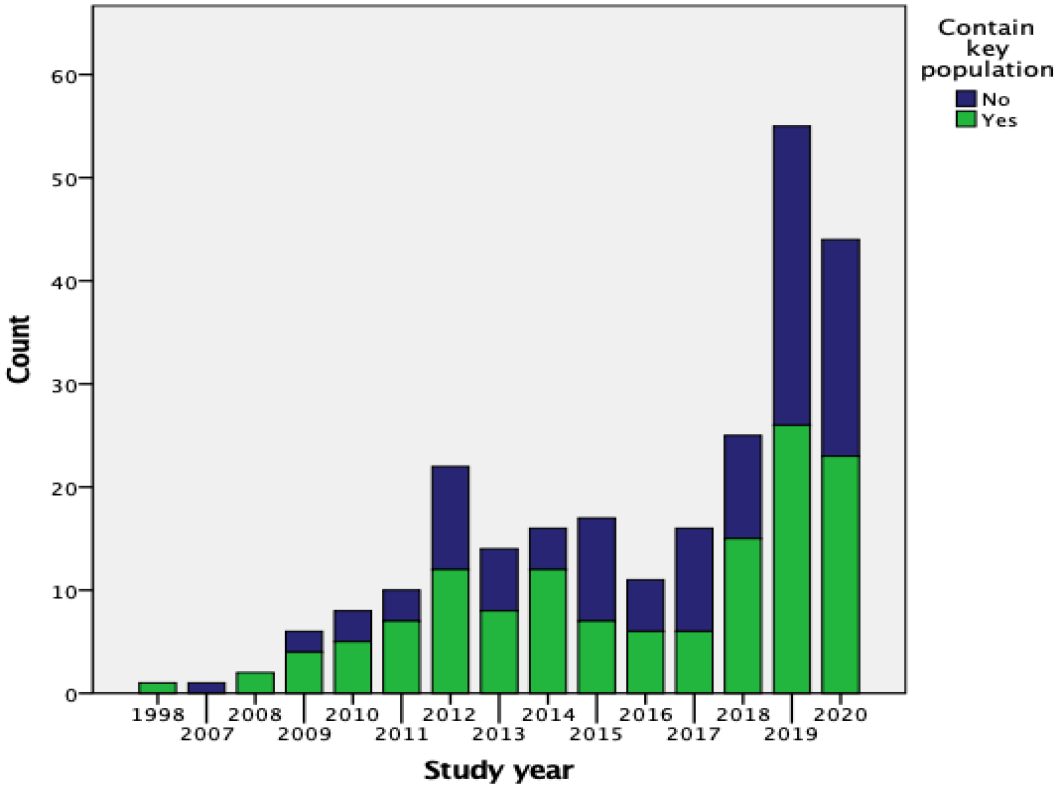

Figure 3: Proportion of studies recruiting key populations. The proportion of studies recruiting key populations has been increasing. In 2019, it reached almost half of published pilot studies in the HIV literature. Key populations were identified using the accepted definitions in the HIV literature, UNAIDS, and the WHO.

Authors disclosed their funding source in 86.7% of studies (k=215). Studies were primarily government funded, either through a direct grant from a government

agency, or indirectly by a government backed grant dispensed from a university. Private funding was found in 15.3% (k=38) of studies, with the most prominent funder being the Bill and Melinda Gates Foundation. Finally, only 9% (k=3.6) studies had industry funding.

Investigators mostly employed experimental designs, with randomized controlled trials (RCTs) and single arm experimental studies comprising 43.5% (k=108) and 51.6% (k=128) of the included studies, respectively. Single arm experimental studies were excluded from the randomization category, as randomization was not employed. Observational studies constituted the remaining studies, with there being 9 cohort studies and 3 observational studies employing alternative designs. These other study designs included 2 case series and a repeated cross-sectional study. Authors also often included a qualitative aspect in their studies, with there being an upward trend (Figure 4). Regardless of study design, trialists primarily sought to evaluate non-pharmacological interventions (k=225; 90.7%). These studies primarily focused on interventions developed to increase participant adherence to interventions, assessing the feasibility of switching antiretroviral therapy (ART) frequency (ex. twice versus three times a day), or improving participant mental health. Pharmacological studies (k=21; 8.5%) were primarily related to the administration of novel ART treatments and evaluating their feasibility.

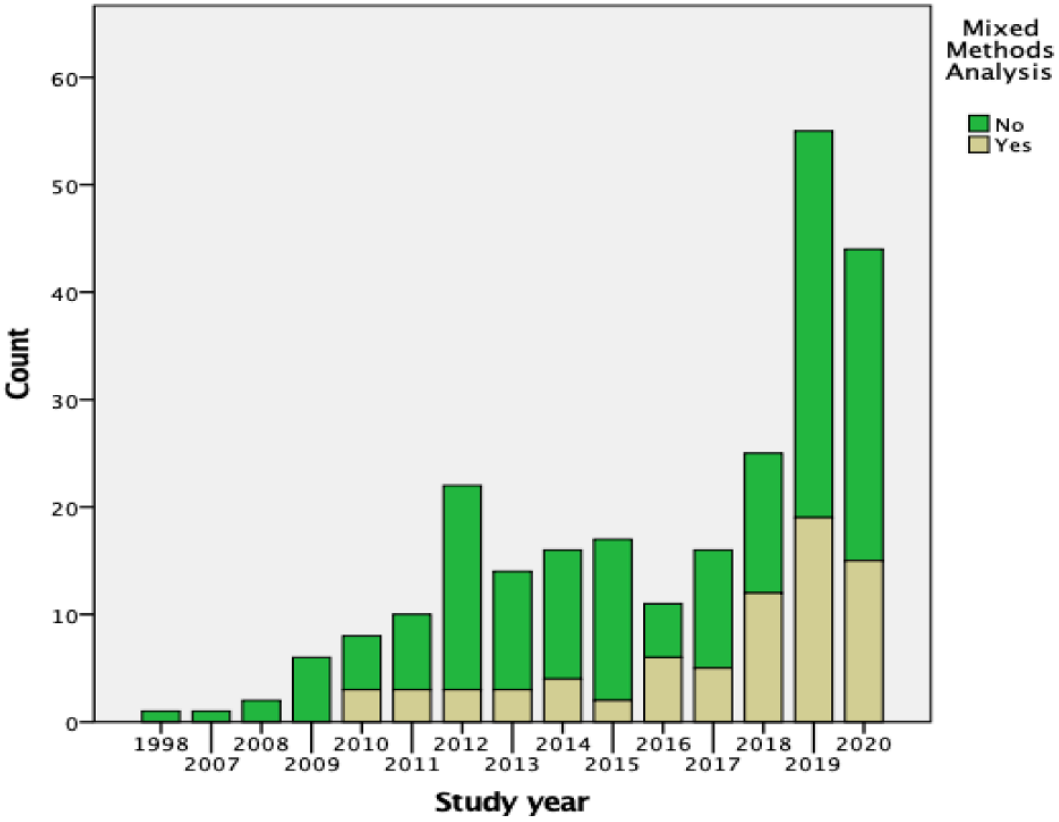

Figure 4: Proportion of studies employing qualitative analysis. The proportion of pilot studies in the HIV literature saw a steady increase until 2018, where close to half of published studies included a qualitative component.

While pilot studies were often conducted in multiple sites (k=105, 42.3%), the majority of studies were conducted at a single site (k=143, 57.7%). However, there was also a geographic disparity in where these studies were conducted. The Americas and Africa were the most common regions (k=137, 55.2%; k=71, 28.6%, respectively). The remaining studies were distributed amongst Southeast Asia, Europe, the Western Pacific or a mix of regions. The disparity was even greater when looking at the income levels of countries where studies were conducted. Most studies were conducted in high income countries, which represented 60.9% of all studies (k=151). This was followed by upper middle income and low-income countries, at 14.9% (k=37) and 13.3% (k=33), respectively. Low- middle income

countries and mixed income countries were least represented, at 10.1% (k=25) and 0.8% (k=2), respectively.

Researchers also recruited a diverse sample amongst the HIV patient population. We found that 134 studies (54%) either included or exclusively recruited key populations in their studies. The complete composition of the patient sample is found in Table 1. ACBs represented the largest subpopulation population amongst studies including key populations at 57% (k=77). Youth (17.2%; K=28), women (17.9%, k=24), and males who have sex with males (MSM) (14.9%, k=20) made up the remaining prominent vulnerable populations present. Other vulnerable populations present were people who inject drugs (PWID), commercial sex workers (CSW), pregnant women, children, and transgender women. A summary of the characteristics of the included studies is reported in Table 1.

Characteristics	Statistic
Design: k (%)	
Randomized control trial	108 (43.5)
Non-randomized trial	140 (56.5)
Intervention Type: k (%)	
Pharmaceutical	21 (8.5)
Non-pharmaceutical	227 (91.5)
Region: k (%)	
Africa	71 (28.6)
Americas	137 (55.2)
Southeast Asia	11 (4.4)
Europe	16 (6.5)
Western Pacific	8 (3.2)
Mixed Region	5 (2)
Country Income Level: k (%)[a]	
High-income	151 (60.9)
Upper-middle-income	37 (14.9)
Lower-middle-income	25 (10.1)
Low-income	33 (13.3)
Mixed- Income	2 (0.8)
Qualitative data: k (%)	75 (30.2)
Key Population: k (%)	
Contain any key population	134 (54)
Specific Key Populations: k (%) (k_{total}=134)	
Injection drug User	4 (1.6)
MSM	20 (8.1)
Incarcerated populations	0 (0)
Commercial sex Workers	1 (0.4)
Pregnant women	11 (4.4)
Children	5 (2)
Youth	28 (11.3)
Indigenous	0 (0)
African, Caribbean, Black	77 (31)
Women	24 (9.1)
Transgender	3 (1.2)
Progression Criteria Prespecified: k (%)	5 (2)
Sample Size: median (min, max)	40 (3, 8794)
Trial Outcome: k (%)	
Proceed to larger study	65 (26.2)
Do not proceed	183(73.8)

MSM= Men who have sex with men
[a]Based upon the categorization by the World Bank(Bank, 2018; WHO, 2020)

Table 1- Baseline Characteristics of Included Studies: k=179

Pilot Study Identification and Criteria

Most studies were easily identifiable as pilot or feasibility studies, with 179 studies (72.2%) including the terms pilot, feasibility, or both in the title. However, in 27 of these studies (15.08%), a feasibility outcome was used to label a study as a pilot or feasibility study. This was often done with the feasibility outcome "acceptance". The remaining 69 studies (27.8%) had no indication of their pilot nature in the title.

Study objectives were often clearly stated in the beginning or as a summarizing statement at the conclusion of the introduction. Authors stated that feasibility would be a primary outcome in 162 studies (65.3%).The remaining studies had goals centered around informing the sample size of the larger study, assessing efficacy, intervention development and to assess the reliability of a measure, with feasibility treated as a secondary outcome. Of the studies included in this review, feasibility outcomes were used as a primary outcome in 157 (63.3%) studies, with the remainder relegating feasibility to be a secondary outcome. Feasibility outcomes and how they were to be assessed was often mentioned in the methods section. However, in 20.16% (k=50) of studies, the feasibility outcomes were only mentioned in the results or discussion exclusively, and not in the methods. Only 5 of the included studies mentioned the progression criteria for their pilot studies.

Authors mentioned their study protocol in 46% of studies (k=114), with 51.75% (k=59) of these providing the registration number or linked to their

published protocol. There was an upward trend in protocol registration over time (Figure 5). Authors of 25.14% of studies (k=45) concluded that moving on to a larger study was feasible, although 33.33% of these studies (k=15) required further modification to proceed. Only 1.7% (k=3) studies claimed that a larger study was not feasible. However, 73.18% (k=131) studies were either ambiguous and did not clearly state whether the study demonstrated feasibility or claimed that the pilot study demonstrated feasibility, but the authors stopped short of recommending progression to the full-scale study.

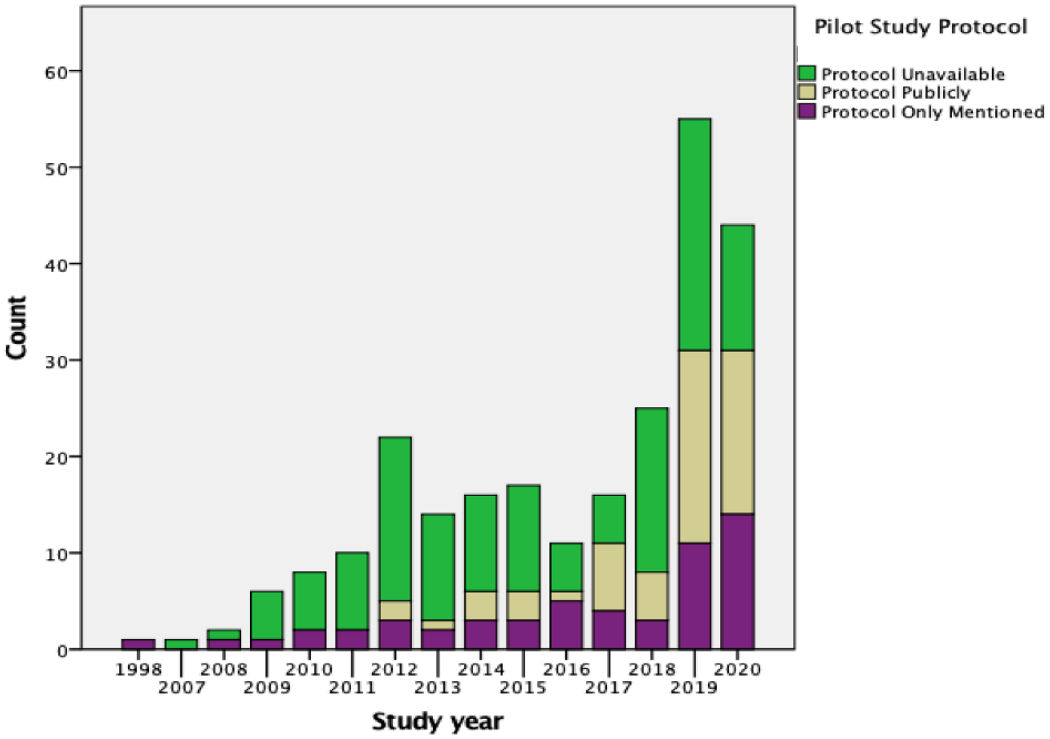

Figure 5: Mention or registration of pilot study protocol. Authors increasingly either made mention of or provided access to their study protocols as time progressed. The proportion of studies doing so never exceeded about 30%.

Feasibility Outcomes and Definitions

We found that the feasibility outcomes assessed in the studies in this review could be grouped into 12 categories in total (Table A1). The most common outcomes were acceptance and retention of participants (k=180, 72.6%; k=135, 54.4%, respectively), as well as evaluating participant enrolment (k=106, 74.6%) and compliance (k=131, 52.8%) to the intervention and study procedures. Trialists also often sought feedback (k=90, 36.3%), primarily from participants and occasionally from study staff.

The definition of different feasibility outcomes varied considerably between studies. The outcome "acceptance" of intervention and study procedures had the greatest variability in its definition, with there being 8 definitions. Acceptance was primarily defined as participant satisfaction with the intervention or study procedures (k=165; 91.7%). Participant satisfaction was assessed either using various satisfaction and acceptance questionnaires (k=56; 31.1%) or participants were simply asked whether they were satisfied with the intervention (k=98; 54.4%). Other feasibility outcomes were also used to define acceptance. This included defining acceptance as participant recruitment or enrolment (k=10; 5.6%), participant retention (k=11; 6.1%), intervention completion (k=16; 8.9%), participant feedback (k=4; 2.2%), and intervention usability (k=2; 1.1%). However, 2 studies (1.1%) did not describe how intervention feedback was defined.

There was also variability in how compliance and participant enrolment were defined. Intervention compliance was defined either as participant attendance

of any intervention sessions, simply adhering to assigned intervention (k= 27; 20.1%) or the number of sessions attended and engagement with the intervention (k= 110; 82.1%). Participant enrolment was primarily defined as the proportion of eligible participants which consented to join the study (k=95; 89.6%). However, 9 studies (8.5%) defined enrolment as the proportion of patients that were recruited and randomized and 2 studies (1.9%) as the defined it as the proportion of participants recruited, consented and were randomized. There was little to no variability in how the remaining feasibility outcomes were defined. Table A1 contains a complete record of the various feasibility outcomes and how they were defined.

Miscellaneous feasibility outcomes were also assessed and defined in 41 studies. The most common of these were implementation (k=34; 13.7%), intervention initiation (k= 6; 2.4%), and consent rate (k=1; 0.4%). Implementation was defined as the ability to deliver the intervention to participants. Initiation was defined as the proportion of eligible participants which were recruited, then consented to join the study and actually began using or were administered the intervention. Finally, consent rate was simply defined as the proportion of eligible patients which consented to joining the study.

Discussion

To the best of our knowledge, this is the first methodological review of pilot studies in the HIV literature. We found that although pilot studies are becoming increasingly common in the HIV literature (Figure 2), there are considerable gaps

in how they are labeled, designed, and how their findings are reported. Additionally, there are inconsistencies in adherence to the pilot study criteria and how the feasibility outcomes are defined. Given that the pilot study field is still in its infancy, these growing pains are to be expected and should serve as a catalyst for the development of additional guidelines.

State of Pilot Studies: Labeling, Pilot Criteria, Design & Reporting

Labeling and Identification. Firstly, there is a need for authors to clearly label their studies and to use the correct terminology when doing so. Making these studies easier to identify is important, as it informs readers of the primary goal of the study and readers understand that any efficacy data included should not be used to inform clinical decisions. We found that most authors (k=179; 72.2%) did label their studies clearly in the title. However, the correct label was not always used. The pilot and feasibility terms were often used interchangeably, or feasibility outcomes, such as acceptance, were used to denote a study's pilot status. As the distinction between the pilot and feasibility terms was only made in 2016, we expect this to improve as this guidance is disseminated and adoption increases (Eldridge, Lancaster, et al., 2016).

It is important to note that using feasibility outcomes to denote the pilot title may make it harder for readers to identify pilot studies. This practice may result in misclassification of these studies, such as readers erroneously extracting information on efficacy to inform medical practice. Other reviews have also found

similar results. In two reviews of pilot studies in the cluster RCT literature and of pilot studies in the *Clinical Rehabilitation* journal over the past 30 years, 83% and 87% of studies contained the terms pilot or feasibility in their title, respectively (Chan, Leyrat, & Eldridge, 2017; Kaur, Figueiredo, Bouchard, Moriello, & Mayo, 2017). However, the latter did find that more than half of the studies used the pilot and feasibility terms interchangeably (Kaur et al., 2017). In addition to these titles being used interchangeably, we also found that several studies used the pilot study title when they were in fact small studies. During the full-text screening stage, we excluded 36 studies containing the pilot or feasibility title for not containing feasibility outcomes at all. These may be studies that failed to recruit their intended sample size or were intended to be small studies and labeled as pilots after the fact. The former may be due to the preference for journals to publish positive trials (Easterbrook, Berlin, Gopalan, & Matthews, 1991; Hopewell, Loudon, Clarke, Oxman, & Dickersin, 2009). However, in the case of smaller studies, these studies should be encouraged and labeled properly as they also have a place in the literature. One study has highlighted how small studies are important in hypothesis testing and in challenging widely held beliefs and common practices (Sackett & Cook, 1993).

Pilot Study Criteria- *Feasibility as primary outcome*. In order to be considered a pilot study, the primary objective of the study has to be to assess the feasibility of a larger study. In fact, the CONSORT extension for pilot studies of RCTs

emphasises this and requires that feasibility outcomes be prespecified and clearly defined (Eldridge, Chan, et al., 2016). We found that authors recognized this and often stated that feasibility was the sole or a co-primary outcome. However, in reality, this was not always the case. In 77 studies (31.1%) assessed, feasibility outcomes were treated as a secondary outcome, even when authors claimed that they were primary outcomes. This is in line with other reviews examining the pilot literature over the past 10 and 30 years, wherein authors switched out the primary outcome (Chan et al., 2017; Kaur et al., 2017; M. Shanyinde, R. M. Pickering, & M. Weatherall, 2011).

However, a bigger issue may be the way in which some studies present their feasibility data. In 50 studies (20.16%), authors reported their feasibility data in the results or discussion section of the paper, while they made no mention of them in any other section. This means that readers are often left guessing whether feasibility data was collected, how the outcomes were defined and what do the findings mean for the larger study. Researchers may miss out on potentially useful feasibility information or mistake them for efficacy studies and use them to inform clinical practice. Once again, we stress the importance of authors being true to the pilot title by having feasibility their primary objective, while ensuring that readers are clearly aware of the outcome(s) of interest and how they are defined, and the data analyzed.

Pilot Study Criteria *-Progression criteria.* An important requirement for pilot studies is the pre-specification of progression criteria (Thabane et al., 2010).

Researchers must outline how they intend to use their findings to inform their future steps. However, only 5 (2%) studies assessed in this review included such criteria. Not reporting progression criteria is problematic, as firstly we are unable to evaluate the quality and appropriateness of the criteria which the author's used to base their decision. Additionally, we are unable to independently evaluate the decision of the investigators with the criteria used to assess the data. We attempted to overcome this by examining the study protocols. We found that 59 studies (23.8%) either provided the registration number for their pilot study or linked the protocol. However, progression criteria were not included in the protocols either. The inability to make this comparison is worrying, as discrepancies between study protocols and publications is a major issue (Chen et al., 2019). Studies with these discrepancies often overestimate their effect size (Chen et al., 2019). It is important to note that other reviews of pilot studies outside the HIV literature report varying levels of progression criteria inclusion in pilot studies. A recent meta-epidemiological review found that only 19.8% of studies included progression criteria (Mbuagbaw et al., 2019), while a review of pilot studies of cluster RCTs found that 89% of studies specified progression criteria (Chan et al., 2017). However, the latter found that only 17% justified and gave reasons for the criteria (Chan et al., 2017). Even when criteria are provided, authors often use vague claims of uncertainty surrounding trial feasibility, cost-effectiveness and participant recruitment (Hallingberg et al., 2018).

The challenges we mentioned are compounded by the fact that the creation of universal feasibility markers is not possible. Each study is unique either in the nature of the intervention, the population of interest, or the setting. As such, researchers may define the same outcomes differently, making it nearly impossible to establish feasibility markers for each outcome. These challenges prevent us from objectively assessing whether the decision made by the authors was appropriate or not. And so, we are left to rely on the expert opinion of the researchers conducting these studies to specify these parameters.

Thus, it is not only important for researchers to craft high quality criteria, but to also make their progression criteria public. While there are no formal guidelines for creating progression criteria, we can provide some guiding principles. Given the subjective nature of how feasibility outcomes and markers of feasibility are defined, we believe that it is important to include input from researchers, potential participants, and research staff. Intra- and inter- group discussions would facilitate the creation of well-rounded and acceptable progression criteria. This collaborative method is based on the Nominal Group Technique (NGT) and allows for the inclusion of different viewpoints to craft progression criteria (Young et al., 2019). It should be noted that since feasibility outcomes are ill-defined, all participants should be provided with a plain-language definition of the relevant concepts at the beginning of this process (Hallingberg et al., 2018).

Additionally, as pilot studies may identify areas of weakness and improvement in study procedures, progression criteria must be flexible to account for this. Instead of a rigid stop-go system, wherein investigators only specify when they will continue or terminate a trial, we recommend a red/amber/green system (Herbert, Julious, & Goodacre, 2019). The amber range would require researchers to specify a middle range wherein the study procedures are amended as per a prespecified method. However, the progression criteria should also include a "rescue plan," to be used in cases where study performance may be less than expected and adjustments are required (Avery et al., 2017).

Currently, the CONSORT extension for pilot studies recommends making the study protocol publicly available, as well as the reporting of protocol violations (Eldridge, Chan, et al., 2016). However, we believe that these reporting requirements should be more explicit, and the inclusion of a requirement for the reporting the progression criteria should be made. Otherwise, authors may deviate from what they prespecified, which has also been found to be a threat to internal validity in full scale trials (Sweetman & Doig, 2011). Additionally, trials that are registered, such as at clinicaltrials.gov, should be required to prespecify and publicly publish their progression criteria. The requirement to register trials was only introduced in 2007,(Zarin, Tse, Williams, & Rajakannan, 2017) and we expect continued refinement to require the inclusion of pilot study specific information, such as progression criteria.

Progression criteria that are clear and visible are vital for improving the decision making, interpretation, and overall quality of pilot studies. Currently, pilot studies are often found to be inconclusive and poorly crafted progression criteria play a major role in this (Arain, Campbell, Cooper, & Lancaster, 2010; Lancaster, Dodd, & Williamson, 2004). However, as we have mentioned before, when pilot studies are well designed, they can help minimize financial and resource waste and may also help in securing funding (Lancaster et al., 2004; Leon, Davis, & Kraemer, 2011; Morgan et al., 2018; van Teijlingen & Hundley, 2002). With the recommendations provided above, researchers will be able to develop better pilot studies, and progression criteria.

Study Design- *Sample size estimations and justifications*. In terms of sample size, we found that estimations and justifications for the pilot studies assessed to be severely lacking. We found that only 23.8% (k=59) of studies had estimations for their sample size. However, even amongst studies that provided a sample size estimation, 28.8% (k=17) of studies did not provide justification for these estimations. Similarly, Chan et al. found that only 44% of pilot cluster RCTs contained justification for their sample size (Chan et al., 2017). As pilot studies do not aim to evaluate the superiority of an intervention over another, it is not necessary to use formal power considerations to inform the sample size of a pilot study. However, it is still necessary to justify the sample size selected.

With the pilot study field still in its infancy, there are currently no hard and fast rules on sample size estimation. However, there are several proposals seeking to provide guidance. Suggestions for sample size determination generally fall into three categories. These include taking a percentage of the larger study's sample size (9% and 50% of larger trial), (Cocks & Torgerson, 2013) having a set minimum number of participants (10-12, 20, or 55 participants per arm) (Cocks & Torgerson, 2013)(Teare et al., 2014) or a using stepped approach determined by standardised effect sizes (k=75, 25, 15 and 10 for standardised effect sizes of ≤0.1, 0.2, 0.5 or 0.8, respectively) (Whitehead, Julious, Cooper, & Campbell, 2015).

Additionally, there have been efforts to create sample size equations specifically for pilot studies. Viechtbauer et al. propose an equation using the probability of a specific problem occurring during the trial to determine a sample size (Viechtbauer et al., 2015). It is important to note that with this approach, very rare events will result in a larger sample size. However, potentially catastrophic worst-case scenario events may be too rare to capture and should be dealt with on an as-needed basis. In cases where multiple problems are being evaluated, the largest sample calculation should be taken (Viechtbauer et al., 2015). With a focus on sequential multiple assignment randomization trial (SMART) pilot studies, Kim et al. instead attempt to calculate the smallest required total sample size (N) using a predetermined minimum number of participants per arm and the expected non-response rate (Kim, 2016).

While we have presented several methods to determine the sample size of pilot studies, it is important to note that no one method is best. This is also reflected in the CONSORT extension for pilot studies, wherein simply a rationale for the numbers included is needed (Eldridge, Chan, et al., 2016). This is in contrast to the formal sample size calculation required for full RCTs (Moher et al., 2010). As we have seen with crafting progression criteria, investigators must use their expertise to apply the current guidance to their pilot studies. This is the same when it comes to sample size calculations as well. We recommend that investigators evaluate their studies to determine which method is best for them, and to be clear with their choice and reasoning. While we found that studies seldomly provide estimations for their sample size and even fewer justify their estimates, we expect this to improve.

Study Design- *Participants in HIV Pilot Studies.* We found that the number of pilot studies has increased with time (Figure 2), meaning that more HIV patients need to be recruited for these studies. It is also important to note that the proportion of HIV patients belonging to a vulnerable population group in these studies has also steadily increased (Figure 3). Given the challenges in recruiting HIV patients, especially those belonging to a vulnerable population group (Joyce K. Anastasi et al., 2005; J. K. Anastasi et al., 2005; Batista et al., 2016; Cook et al., 2018; De La Rosa et al., 2012; el-Sadr & Capps, 1992; Fortune et al., 2010; Gemmill et al., 2012; Hoffman et al., 2019; Loutfy et al., 2014; Menezes et al., 2011; Nalubega & Evans, 2015; Silvestre et al., 2006; Yehia et al., 2015) , researchers have begun to recognize

the need to employ novel recruitment and retention methods. Pilot studies are important to evaluate these methods to ensure a robust methodological approach in the subsequent trial. With this being said, the pilot phase is where researchers should be daring in implementing new methods to recruit participants.

Currently, there is some guidance on how to better reach disadvantaged groups in full-scale trials (Bonevski et al., 2014). While these solutions are not specific to HIV patients and pilot studies, they are still useful and should be incorporated into study procedures. Firstly, it is important to reduce barriers to participation. This begins with developing and strengthening bonds with community organizations (Bonevski et al., 2014). This not only involves knowing the people in the community, but also including them in the study design process. Digital tools, such as social media or email can also help reach patients unable to publicly self-identify due to social stigma or for being left out by those gate-keeping care (Bonevski et al., 2014). This knowledge should be used to adjust data collection methods and tools by incorporating inclusive language and making their administration flexible. Given that some may be externally motivated, incentives should be provided (Bonevski et al., 2014). Pilot studies should be used to evaluate which incentives work best. While RCTs benefit from randomization, which helps limit bias, they also require rigid and generally nonpragmatic methods. Therefore, investigators may benefit from employing study designs other than the classic RCT (Bonevski et al., 2014). It should be noted that these steps will require additional resources researchers and lengthen the timeline. Strengthening community bonds

will require time and incorporating flexible procedures and new data collection tools may require added personnel, equipment and training.

Just as with progression criteria, we also suggest that participants be involved in creating these procedures. By incorporating innovative design features, strengthening community bonds and employing participant centered approaches in pilot studies, study quality will undoubtedly improve, and recruitment and retention targets will be easier to meet. In addition, recruitment targets should be based on rates per unit of time (ex. per month) and not an absolute number by a specific date (Avery et al., 2017).

Study Design- *Qualitative methods in pilot studies.* We found that investigators are increasingly employing participant centric outcomes and qualitative methods. We believe that the increasing role of such methodology is important in improving the quality of pilot studies. There have been recent calls for greater inclusion of qualitative methods in pilot studies (Bertram, Moore, Wylde, & Gooberman-Hill, 2019; Hallingberg et al., 2018). Looking at our results, we believe that researchers are beginning to realize the utility of employing qualitative methods. In our list of feasibility outcomes (Table A1), we see that the most common outcome, acceptance (k=180 studies), was participant centric. In addition to this being the most common outcome, we also found that there was great variability in how this was defined and evaluated, with there being 8 definitions and two main ways of evaluating it (specific acceptance tool or freeform, open-ended question about acceptance).

Researchers also often elicited feedback from participants in an unstructured and freeform manner (k=90; 36.3%). This heterogeneity in definition and measurement demonstrates the emphasis investigators place on what patients have to say about the intervention and study procedures. Not only did investigators seek to assess how best to recruit and retain participants, but they also sought to learn about participant satisfaction with the intervention and study procedures directly. In addition, looking at Figure 5, we see that the use of qualitative methods in pilot studies in the HIV literature is trending upwards. This trend is in line with other published evidence, which found that between 2008 and 2010, there was a 28% increase in the use of qualitative methods in randomized feasibility studies (O'Cathain, Thomas, Drabble, Rudolph, & Hewison, 2013). The use of patient centered outcomes and qualitative methods has been shown to help refine study procedures, including optimizing recruitment and retention (Donovan et al., 2016; Elliott, Husbands, Hamdy, Holmberg, & Donovan, 2017). By incorporating these methods in pilot studies, investigators are able to set realistic targets, craft pragmatic procedures and ask and answer a wider range of questions, while gaining granular detail (Bertram et al., 2019; O'Cathain et al., 2015). In addition, the involvement of participants in the design process of the study procedures and intervention motivates participant interest in the study. The increasing prevalence of qualitative methodology in pilot studies is encouraging.

While pilot studies are meant to be smaller versions of a larger study, authors may opt to deviate from being an exact replica and incorporate qualitative

aspects into their pilot to gain deeper insight into how to improve their study design. Qualitative methods may be incorporated in any aspect of a pilot study, this includes the research question, study design, analysis and reporting, and even the composition of the research team (O'Cathain et al., 2015). While discussing progression criteria and sample size calculations above, we advocate for methods that allow for flexibility. Qualitative methods promote this, as researchers gain a greater insight during the pilot stage of added attention and flexibility is needed. Qualitative methods not only have a place in full scale studies, but also in exploring their uncertainties and optimising the intervention or trial procedures during the pilot stage.

Reporting of HIV Pilot Studies- The CONSORT extension for pilot studies is a provides comprehensive guidance on how to report findings from pilot studies and should be used as the basis of reporting by researchers (Eldridge, Chan, et al., 2016). However, additional emphasis is required to address the reporting shortcomings we have found in the HIV literature. Studies in the HIV field require pilot testing as they involve a population that is relatively challenging to identify, recruit and retain (J. K. Anastasi et al., 2005; Cook et al., 2018; De La Rosa et al., 2012; el-Sadr & Capps, 1992; Fortune et al., 2010; Hoffman et al., 2019; Menezes et al., 2011; Nalubega & Evans, 2015; Silvestre et al., 2006; Yehia et al., 2015). Working with such a population requires researchers to tailor their approach. Given the community, recruitment methods may have to differ, interventions may have to be

adjusted, and data collection tools may need to be tailored to be culturally sensitive. Improving reporting in the HIV pilot literature begins with clearly labeling pilot studies using the correct title. Using the correct label and making it clear as this informs readers to expect feasibility related information as the primary focus. Therefore, it is important for researchers to follow up on this by clearly reporting and defined the feasibility outcomes. In addition, the role of the feasibility outcome(s) should also be outlined (if there are multiple feasibility outcomes, what is the primary? Or are there co-primary outcomes?). It is also important to report the intended recruitment strategy and sample size and to justify latter. Finally, researchers must outline the progression criteria by which they will evaluate their data. This should be prespecified and reported in the study protocol, which should be made public. Authors should also make it clear that any efficacy data reported should be taken with a grain of salt as pilot studies are not powered to draw such conclusions. While the CONSORT extension is intended for randomized studies, we believe that authors conducting observational studies can use the reporting guide while slightly modifying it to account for the lack of randomization and the more pragmatic nature of these studies.

Defining Feasibility Outcomes

When studies claimed to evaluate the same feasibility outcome, we found that they are often defined differently. We saw this with common outcomes such as acceptance, compliance and participant enrollment. We found that acceptance alone

had 8 definitions. However, even though most investigators defined acceptance as participant satisfaction (k=165; 91.7%), this was captured using various satisfactions instruments (k=56; 31.1%) or participants were simply asked what they thought of the intervention or trial procedures to measure satisfaction (k=98; 54.4%). The former requires the measures used to be reliable and valid. The latter relies on a participant's perception of what is satisfaction or for research staff to record and interpret free-form answers. We should also note that other researchers have found feasibility outcomes we did not, such as "selection of most appropriate primary outcome measure,"(Lancaster et al., 2004). It is clear that concepts commonly believed to be universal are not.

With this being said, we believe that instead of striving to capture and define each and every feasibility outcome, outcomes should instead be organized in broad categories. Using these categories as a guide, researchers should then clearly state and define their outcome(s) of interest, including how they hope to assess and evaluate this feasibility data. Researchers require this flexibility to accurately capture exactly what they are interested in. Currently, there are two suggestions of how feasibility outcomes should be categorized. One proposal includes 7 areas of focus and these are 1) acceptance, 2) implementation, 3) practicality, 4) adaptation, 5) implementation 6) integration, 7) expansion, 8) limited efficacy (Bowen et al., 2009). Given that these categories are based on common feasibility outcomes, they are intuitive and easy to use. However, being based on feasibility outcomes also means that some outcomes may not fit neatly within these categories. This may lead

to future confusion and fragmentation when the need arises to expand these categories. An alternative categorization is based on the 4 reasons to conduct a pilot study. These categories are 1) assessing study processes, 2) study resources evaluation, 3) management of potential human and data management problems, and 4) assessment of intervention safety, dose response, and variance of effect (Thabane et al., 2010). We believe the latter categories are more appropriate as they are sufficiently broad, while still being specific to pilot studies. This broadness allows researchers the flexibility needed to tailor their outcomes to their unique cases, which is something we have emphasised continuously.

Pilot studies, by their highly specific nature, promote heterogeneity in the methods employed and the way feasibility outcomes are defined and assessed. While it may be tempting to standardize pilot studies, we believe that this flexibility should be promoted. However, there is a need for greater emphasis for better reporting of outcomes and how they are defined. This way, readers understand what researchers are interested in assessing, instead of relying on their own conceptions of an outcome.

Designing Better Pilot Studies

While we did not evaluate which factors influenced an author's decision to progress with the larger trial, from our results, we can provide general advice when designing a pilot study. When designing pilot studies, we advise researchers to take a collaborative approach while also incorporating qualitative methods to gain

greater insight from other stakeholders. This is especially important when defining feasibility outcomes and crafting progression criteria and setting study procedures. It is also important for researchers to be transparent, upfront and clear with the study objectives and how decisions are made. While these are not guidelines on how to develop a pilot study, they are guiding principles to keep in mind throughout the process. As the pilot study field matures, we expect the current reporting guidance, such as the CONSORT extension, to be more widely adopted and for additional detailed guidelines to be developed as well.

While we are unable to provide design recommendations, Charlesworth et al. have created a checklist, based on the CONSORT 2010 statement, to help investigators determine whether data from their pilot studies can be included with that of the full trial (Charlesworth, Burnell, Hoe, Orrell, & Russell, 2013). This decision is based on how similar the pilot study methods are to the those of the full trial. While this does not directly guide researchers on how to design their pilot studies, it does reinforce the need for these studies to be representative of the larger study when designing the pilot. Charlesworth et al. state that in cases where only minor to moderate changes are needed, the data is easily incorporated (Charlesworth et al., 2013). Otherwise it is recommended that researchers return to the pilot stage and not to incorporate the data (Charlesworth et al., 2013). However, it is important to note that these decisions are entirely up to the investigators.

Weaknesses, Strengths, and Future Directions

Our meta-epidemiological study has some weaknesses. Firstly, we were reliant on the authors' conclusions to determine whether the study was indeed feasible and should proceed to the larger study. While ideally, we would prefer to have a standardized method to allow us to determine this, such guidance would be difficult to provide. Therefore, we are forced to rely on the judgement of the researchers.

While our study does have weaknesses, it also has several strengths. Our review was robust as our search was highly sensitive, as demonstrated by the exceptionally large number of studies screened (k=9,297). Our search was conducted in Medline, Embase, and CENTRAL, with no restrictions. The concepts searched, pilot studies and HIV, were purposefully broad as we are interested in all HIV interventions. In addition, with our study being specific to the HIV literature, we are able to evaluate a particular area of research where pilot studies will increasingly play a bigger role.

Future studies could expand on this work and evaluate how data quality impacts the decision researchers make at the end of their pilot. While formal guidance on assessing the quality of pilot studies does not yet exist yet, current tools could be used, given that pilot studies are simply scaled down versions of their main studies. Therefore, authors may use the Cochrane Risk of Bias tool for RCTs and ROBINS-I or the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) for observational studies (Sterne et al., 2016),(von Elm et al., 2014).

Conclusion

Pilot studies are increasingly being used in the HIV field. However, there are many gaps in the reporting of pilot studies in the HIV literature. We found that assessing feasibility was not always the primary goal of the pilot studies evaluated. In addition, many major decisions, such as selecting a sample size, crafting progression criteria and defining feasibility outcomes were not frequently done or are inconsistent in how they are done. As research in HIV populations is resource intensive, pilot studies will play a larger role in this area, given the ability of pilot studies to determine whether a larger study would be feasible or not. And so higher quality pilot studies are needed. Our project was the first step in this process. By evaluating the current state of pilot studies in the HIV literature, we have highlighted the need for additional guidance on how to design and report on pilot studies in HIV research. In addition to guidance, we suggest that there must be stronger reporting requirements for studies registered on platforms such as clinicaltrials.gov. An improvement in HIV pilot studies may consequently lead to more efficient trials and potential cost-savings.

Appendix

Long Table

A1- Feasibility outcomes and definitions

Outcome[a] (No., %)	Definition
Acceptance (k=180, 72.6%)	a) Participant satisfaction with the intervention and study procedures (k=165; 91.7%). b) Participant recruitment or enrolment rate (k=10; 5.6%). c) Intervention completion (k=16; 8.9%) d) Participant retention (k=11; 6.1%). e) Participant feedback (k=4; 2.2%). f) Intervention usability (k=2; 1.1%).
Contamination (k=2, 0.8%)	The proportion of participants that deviated from their allocated intervention and partook in the alternative intervention.
Compliance[b] (k=131, 52.8%)	a) Participant attendance of any intervention sessions or simply adherence to assigned intervention (k=27; 20.1%). b) The number of sessions attended and engagement with the intervention (k=110; 82.1%).
Data Completion (k=9, 5%)	The proportion of data expected to be collected which was not.
Enrolment (k=106, 42.7%)	a) The proportion of eligible participants which consented to join the study (k=95; 90.5%). b) The proportion of participants which were recruited and randomized (k=9; 8.5%).

	c) The proportion of participants recruited, consented and were randomized (k=2. 1.9%)
Feedback (k=90, 36.3%)	What participants and study staff thought of the intervention, study procedures and their time in the trial.
Fidelity (k=37, 14.9%)	The ability of the trialists to adhere to study protocol.
Randomization (k=11, 4.4%)	The ability to successfully randomly allocate participants to the different arms in a trial.
Retention (k=135, 54.4%)	The proportion of participants which remained in the study till the primary endpoint (either end of the intervention or a set follow-up period).
Resources (k=16, 6.5%)	An evaluation of the resources required to conduct the study.
Timeliness of Intervention (k=2, 0.8%)	Assessment of the ability to administer the intervention in the prespecified time.
Other (k=34, 13.7%)	a) Implementation- The ability to deliver the intervention to participants (k=17; 65.38%). b) Initiation- The proportion of eligible participants which were recruited, consented to join the study and actually began using or were administered the intervention (k=6; 23.07%). c) Consent rate- The proportion of eligible patients which consented to joining the study (k=1; 3.8%).

[a] Studies may possess several feasibility outcomes
[b] Participant acceptance was often defined and assessed in multiple ways in the same study

References:

Anastasi, J. K., Capili, B., Kim, G. H., & Chung, A. (2005). Clinical trial recruitment and retention of a vulnerable population: Hiv patients with chronic diarrhea. *Gastroenterology Nursing*. doi:10.1097/00001610-200511000-00002

Anastasi, J. K., Capili, B., Kim, G. H., & Chung, A. (2005). Clinical trial recruitment and retention of a vulnerable population: HIV patients with chronic diarrhea. *Gastroenterol Nurs, 28*(6), 463-468. doi:10.1097/00001610-200511000-00002

Arain, M., Campbell, M. J., Cooper, C. L., & Lancaster, G. A. (2010). What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Med Res Methodol, 10*, 67. doi:10.1186/1471-2288-10-67

Asiimwe, S. B., Kanyesigye, M., Bwana, B., Okello, S., & Muyindike, W. (2016). Predictors of dropout from care among HIV-infected patients initiating antiretroviral therapy at a public sector HIV treatment clinic in sub-Saharan Africa. *BMC Infectious Diseases*. doi:10.1186/s12879-016-1392-7

Avery, K. N., Williamson, P. R., Gamble, C., O'Connell Francischetto, E., Metcalfe, C., Davidson, P., . . . members of the Internal Pilot Trials Workshop supported by the Hubs for Trials Methodology, R. (2017). Informing efficient randomised controlled trials: exploration of challenges in developing progression criteria for internal pilot studies. *BMJ Open, 7*(2), e013537. doi:10.1136/bmjopen-2016-013537

Bank, T. W. (2018). World Bank Open Data. *World Bank*. doi:10.1145/2750858.2804284

Batista, P., Deren, S., Banfield, A., Silva, E., Cruz, M., Garnes, P., . . . Markowitz, M. (2016). Challenges in Recruiting People Who Use Drugs for HIV-Related Biomedical Research: Perspectives from the Field. *AIDS Patient Care and STDs*. doi:10.1089/apc.2016.0135

Bekker, L. G., Johnson, L., Wallace, M., & Hosek, S. (2015). Building our youth for the future. *J Int AIDS Soc, 18*(2 Suppl 1), 20027. doi:10.7448/IAS.18.2.20027

Bertram, W., Moore, A., Wylde, V., & Gooberman-Hill, R. (2019). Optimising recruitment into trials using an internal pilot. *Trials, 20*(1), 207. doi:10.1186/s13063-019-3296-5

Bonevski, B., Randell, M., Paul, C., Chapman, K., Twyman, L., Bryant, J., . . . Hughes, C. (2014). Reaching the hard-to-reach: A systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC Medical Research Methodology*. doi:10.1186/1471-2288-14-42

How We Design Feasibility Studies, (2009).

Bugge, C., Williams, B., Hagen, S., Logan, J., Glazener, C., Pringle, S., & Sinclair, L. (2013). A process for Decision-making after Pilot and

feasibility Trials (ADePT): Development following a feasibility study of a complex intervention for pelvic organ prolapse. *Trials*. doi:10.1186/1745-6215-14-353

Predictors of Adult Retention in HIV Care: A Systematic Review, (2018).

Avoidable waste in the production and reporting of research evidence, (2009).

Chan, C. L., Leyrat, C., & Eldridge, S. M. (2017). Quality of reporting of pilot and feasibility cluster randomised trials: a systematic review. *BMJ Open, 7*(11), e016970. doi:10.1136/bmjopen-2017-016970

Charlesworth, G., Burnell, K., Hoe, J., Orrell, M., & Russell, I. (2013). Acceptance checklist for clinical effectiveness pilot trials: a systematic approach. *BMC Med Res Methodol, 13*, 78. doi:10.1186/1471-2288-13-78

Chen, T., Li, C., Qin, R., Wang, Y., Yu, D., Dodd, J., . . . Cornelius, V. (2019). Comparison of Clinical Trial Changes in Primary Outcome and Reported Intervention Effect Size Between Trial Registration and Publication. *JAMA Netw Open, 2*(7), e197242. doi:10.1001/jamanetworkopen.2019.7242

Cocks, K., & Torgerson, D. J. (2013). Sample size calculations for pilot randomized trials: a confidence interval approach. *J Clin Epidemiol, 66*(2), 197-201. doi:10.1016/j.jclinepi.2012.09.002

Cook, C., Mack, J., & Cottler, L. B. (2018). Research participation, trust, and fair compensation among people living with and without HIV in Florida. *AIDS Care, 30*(1), 27-31. doi:10.1080/09540121.2017.1338656

De La Rosa, M., Babino, R., Rosario, A., Martinez, N. V., & Aijaz, L. (2012). Challenges and strategies in recruiting, interviewing, and retaining recent Latino immigrants in substance abuse and HIV epidemiologic studies. *Am J Addict, 21*(1), 11-22. doi:10.1111/j.1521-0391.2011.00193.x

Djomand, G., Quaye, S., & Sullivan, P. S. (2014). HIV epidemic among key populations in west Africa. *Curr Opin HIV AIDS, 9*(5), 506-513. doi:10.1097/COH.0000000000000090

Donovan, J. L., Rooshenas, L., Jepson, M., Elliott, D., Wade, J., Avery, K., . . . Blazeby, J. M. (2016). Optimising recruitment and informed consent in randomised controlled trials: the development and implementation of the Quintet Recruitment Intervention (QRI). *Trials, 17*(1), 283. doi:10.1186/s13063-016-1391-4

Easterbrook, P. J., Berlin, J. A., Gopalan, R., & Matthews, D. R. (1991). Publication bias in clinical research. *Lancet, 337*(8746), 867-872. doi:10.1016/0140-6736(91)90201-y

el-Sadr, W., & Capps, L. (1992). The challenge of minority recruitment in clinical trials for AIDS. *JAMA, 267*(7), 954-957. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1734108

Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., . . . Tugwell, P. (2016). CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. *Pilot and Feasibility Studies*. doi:10.1186/s40814-016-0105-8

Eldridge, S. M., Lancaster, G. A., Campbell, M. J., Thabane, L., Hopewell, S., Coleman, C. L., & Bond, C. M. (2016). Defining feasibility and pilot studies in preparation for randomised controlled trials: Development of a conceptual framework. *PLoS ONE*. doi:10.1371/journal.pone.0150205

Elliott, D., Husbands, S., Hamdy, F. C., Holmberg, L., & Donovan, J. L. (2017). Understanding and Improving Recruitment to Randomised Controlled Trials: Qualitative Research Approaches. *Eur Urol, 72*(5), 789-798. doi:10.1016/j.eururo.2017.04.036

Fortune, T., Wright, E., Juzang, I., & Bull, S. (2010). Recruitment, enrollment and retention of young black men for HIV prevention research: experiences from The 411 for Safe Text project. *Contemp Clin Trials, 31*(2), 151-156. doi:10.1016/j.cct.2009.12.004

Gemmill, R., Williams, A. C., Cooke, L., & Grant, M. (2012). Challenges and strategies for recruitment and retention of vulnerable research participants: Promoting the benefits of participation. *Applied Nursing Research*. doi:10.1016/j.apnr.2010.02.003

The Role of Open Access in Reducing Waste in Medical Research, (2014).

Hallingberg, B., Turley, R., Segrott, J., Wight, D., Craig, P., Moore, L., . . . Moore, G. (2018). Exploratory studies to decide whether and how to proceed with full-scale evaluations of public health interventions: a systematic review of guidance. *Pilot Feasibility Stud, 4*, 104. doi:10.1186/s40814-018-0290-8

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*. doi:10.1016/j.jbi.2008.08.010

Herbert, E., Julious, S. A., & Goodacre, S. (2019). Progression criteria in trials with an internal pilot: an audit of publicly funded randomised controlled trials. *Trials, 20*(1), 493. doi:10.1186/s13063-019-3578-y

Hoffman, K. A., Baker, R., Kunkel, L. E., Waddell, E. N., Lum, P. J., McCarty, D., & Korthuis, P. T. (2019). Barriers and facilitators to recruitment and enrollment of HIV-infected individuals with opioid use disorder in a clinical trial. *BMC Health Serv Res, 19*(1), 862. doi:10.1186/s12913-019-4721-x

Hopewell, S., Loudon, K., Clarke, M. J., Oxman, A. D., & Dickersin, K. (2009). Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev*(1), MR000006. doi:10.1002/14651858.MR000006.pub3

Hupe, M. (2019). EndNote X9. *Journal of Electronic Resources in Medical Libraries*. doi:10.1080/15424065.2019.1691963

Irving, S. Y., & Curley, M. A. Q. (2008). Challenges to conducting multicenter clinical research ten points to consider. *AACN Advanced Critical Care*. doi:10.1097/01.AACN.0000318119.67061.0f

Kaur, N., Figueiredo, S., Bouchard, V., Moriello, C., & Mayo, N. (2017). Where have all the pilot studies gone? A follow-up on 30 years of pilot studies in Clinical Rehabilitation. *Clin Rehabil, 31*(9), 1238-1248. doi:10.1177/0269215517692129

Kellermeyer, L., Harnke, B., & Knight, S. (2018). Covidence. *Journal of the Medical Library Association*. doi:10.5195/jmla.2018.513

Kim, H. (2016). A Sample Size Calculator for SMART Pilot Studies. *SIAM Undergraduate Research Online*. doi:10.1137/15s014058

Lancaster, G. A., Dodd, S., & Williamson, P. R. (2004). Design and analysis of pilot studies: recommendations for good practice. *J Eval Clin Pract, 10*(2), 307-312. doi:10.1111/j..2002.384.doc.x

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *J Psychiatr Res, 45*(5), 626-629. doi:10.1016/j.jpsychires.2010.10.008

Loutfy, M. R., V, L. K., Mohammed, S., Wu, W., Muchenje, M., Masinde, K., . . . Tharao, W. (2014). Recruitment of HIV-Positive Women in Research: Discussing Barriers, Facilitators, and Research Personnel's Knowledge. *The Open AIDS Journal*. doi:10.2174/1874613601408010058

Macdonald, V., Verster, A., & Baggaley, R. (2017). A call for differentiated approaches to delivering HIV services to key populations. *J Int AIDS Soc, 20*(Suppl 4), 21658. doi:10.7448/IAS.20.5.21658

Mbuagbaw, L., Kosa, S. D., Lawson, D. O., Stalteri, R., Olaiya, O. R., Alotaibi, A., & Thabane, L. (2019). The reporting of progression criteria in protocols of pilot trials designed to assess the feasibility of main trials is insufficient: A meta-epidemiological study. *Pilot and Feasibility Studies*. doi:10.1186/s40814-019-0500-z

Menezes, P., Eron, J. J., Jr., Leone, P. A., Adimora, A. A., Wohl, D. A., & Miller, W. C. (2011). Recruitment of HIV/AIDS treatment-naive patients to clinical trials in the highly active antiretroviral therapy era: influence of gender, sexual orientation and race. *HIV Med, 12*(3), 183-191. doi:10.1111/j.1468-1293.2010.00867.x

Implementing clinical trials on an international platform: Challenges and perspectives, (2012).

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gotzsche, P. C., Devereaux, P. J., . . . Consolidated Standards of Reporting Trials, G. (2010). CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol, 63*(8), e1-37. doi:10.1016/j.jclinepi.2010.03.004

Morgan, B., Hejdenberg, J., Hinrichs-Krapels, S., & Armstrong, D. (2018). Do feasibility studies contribute to, or avoid, waste in research? *PLoS ONE*. doi:10.1371/journal.pone.0195951

Murad, M. H., & Wang, Z. (2017). Guidelines for reporting meta-epidemiological methodology research. *Evidence-Based Medicine*. doi:10.1136/ebmed-2017-110713

Nalubega, S., & Evans, C. (2015). Participant views and experiences of participating in HIV research in sub-Saharan Africa: a qualitative systematic review. *JBI Database System Rev Implement Rep, 13*(5), 330-420. doi:10.11124/jbisrir-2015-2051

O'Cathain, A., Hoddinott, P., Lewin, S., Thomas, K. J., Young, B., Adamson, J., . . . Donovan, J. L. (2015). Maximising the impact of qualitative research in feasibility studies for randomised controlled trials: guidance for researchers. *Pilot Feasibility Stud, 1*, 32. doi:10.1186/s40814-015-0026-y

O'Cathain, A., Thomas, K. J., Drabble, S. J., Rudolph, A., & Hewison, J. (2013). What can qualitative research do for randomised controlled trials? A systematic mapping review. *BMJ Open, 3*(6). doi:10.1136/bmjopen-2013-002889

Rao, A., Stahlman, S., Hargreaves, J., Weir, S., Edwards, J., Rice, B., . . . Baral, S. (2017). Sampling Key Populations for HIV Surveillance: Results From Eight Cross-Sectional Studies Using Respondent-Driven Sampling and Venue-Based Snowball Sampling. *JMIR Public Health Surveill, 3*(4), e72. doi:10.2196/publichealth.8116

Sackett, D. L., & Cook, D. J. (1993). Can we learn anything from small trials? *Ann N Y Acad Sci, 703*, 25-31; discussion 31-22. doi:10.1111/j.1749-6632.1993.tb26331.x

Shanyinde, M., Pickering, R. M., & Weatherall, M. (2011). Questions asked and answered in pilot and feasibility randomized controlled trials. *BMC Medical Research Methodology*. doi:10.1186/1471-2288-11-117

Shanyinde, M., Pickering, R. M., & Weatherall, M. (2011). Questions asked and answered in pilot and feasibility randomized controlled trials. *BMC Med Res Methodol, 11*, 117. doi:10.1186/1471-2288-11-117

Silvestre, A. J., Hylton, J. B., Johnson, L. M., Houston, C., Witt, M., Jacobson, L., & Ostrow, D. (2006). Recruiting minority men who have sex with men for HIV research: results from a 4-city campaign. *Am J Public Health, 96*(6), 1020-1027. doi:10.2105/AJPH.2005.072801

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., . . . Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ (Online)*. doi:10.1136/bmj.i4919

Sweetman, E. A., & Doig, G. S. (2011). Failure to report protocol violations in clinical trials: A threat to internal validity? *Trials*. doi:10.1186/1745-6215-12-214

Teare, M. D., Dimairo, M., Shephard, N., Hayman, A., Whitehead, A., & Walters, S. J. (2014). Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. *Trials, 15*, 264. doi:10.1186/1745-6215-15-264

A tutorial on pilot studies: The what, why and how, (2010).

van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies. *Nurs Stand, 16*(40), 33-36. doi:10.7748/ns2002.06.16.40.33.c3214

Viechtbauer, W., Smits, L., Kotz, D., Bude, L., Spigt, M., Serroyen, J., & Crutzen, R. (2015). A simple formula for the calculation of sample size in pilot studies. *J Clin Epidemiol, 68*(11), 1375-1379. doi:10.1016/j.jclinepi.2015.04.014

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2014). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *International Journal of Surgery*. doi:10.1016/j.ijsu.2014.07.013

Whitehead, A. L., Julious, S. A., Cooper, C. L., & Campbell, M. J. (2015). Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Statistical Methods in Medical Research*. doi:10.1177/0962280215588241

WHO. (2020). Definition of regional groupings. Retrieved from https://www.who.int/healthinfo/global_burden_disease/definition_regions/en/

Yehia, B. R., Stewart, L., Momplaisir, F., Mody, A., Holtzman, C. W., Jacobs, L. M., . . . Shea, J. A. (2015). Barriers and facilitators to patient retention in HIV care. *BMC Infect Dis, 15*, 246. doi:10.1186/s12879-015-0990-0

Young, H. M. L., Goodliffe, S., Madhani, M., Phelps, K., Regen, E., Locke, A., . . . Conroy, S. (2019). Co-producing Progression Criteria for Feasibility Studies: A Partnership between Patient Contributors, Clinicians and Researchers. *Int J Environ Res Public Health, 16*(19). doi:10.3390/ijerph16193756

Zarin, D. A., Tse, T., Williams, R. J., & Rajakannan, T. (2017). Update on Trial Registration 11 Years after the ICMJE Policy Was Established. *N Engl J Med, 376*(4), 383-391. doi:10.1056/NEJMsr1601330