MEETING REPORT

Best practices in analytic approaches to assess the effect of hormonal contraception on HIV acquisition in observational data

January 24th and 25th, Seattle, WA

Jennifer E. Balkus, PhD, MPH

MEETING OBJECTIVE:
To make recommendations for best analytic practices with observational data when assessing the effect of hormonal contraception on HIV acquisition.

Editorial note: Hormonal contraception (HC) includes a number of different pregnancy prevention methods including, but not limited to, progestogen-only injectables such as depot medroxyprogesterone acetate (DMPA) or norethisterone enanthate (NET-EN), combined estrogen and progestin injectables, combined estrogen and progestin oral contraceptive pills (COCs), progestin-only pills (POPs), levonorgestrel-containing implants (e.g. Norplant, Jadelle, Sino-Implant), and levonorgestrel-containing intrauterine devices (e.g. Mirena IUD). Each of the methods listed above may have a different effect, if any, on risk of HIV acquisition. Throughout the report, the term "hormonal contraception" or "HC" is used for editorial convenience, recognizing that the author of this report and meeting participants acknowledge the importance of understanding the safety of each individual method as it relates to HIV acquisition risk.

Introduction

The possibility that use of certain methods of HC may impact the risk of HIV acquisition is of great concern to the global public health community, contraceptive clients and their partners, and providers. To date, epidemiological studies on this subject have used varied analytic approaches and generated heterogeneous results. In this context of uncertainty, issuing clear policy recommendations is complicated, and the need for clearer answers is urgent. While there have been multiple epidemiological studies on HC and HIV acquisition in women, to our knowledge, there has not been a meeting bringing together subject and methods experts to discuss analytical approaches to assessing this complex issue. Additional analyses of observational datasets are currently underway, providing an opportunity to offer analytic guidance. The goal of this meeting was to provide recommendations and guidance concerning best analytic practices, which could be applied to future analyses, and ultimately contribute to the broader goal of clarity around the science, policy, and communications on this important public health issue.

How do findings from observational studies impact WHO recommendations for the medical eligibility criteria for contraceptive use?

Dr. Sharon Phillips from the World Health Organization (WHO) provided an overview of the WHO process for determining the Medical Eligibility Criteria (MEC) for Contraceptive Use. The goal of the MEC for contraceptive use is to provide policy and decision makers, and the scientific community, with recommendations that can be used for developing or revising national guidelines on medical eligibility criteria for contraceptive use. The recommendations are intended to provide a basis to rationalize provision of various contraceptives in view of the most

up-to-date information available. Because national and programmatic contexts vary so greatly, it is inappropriate to set firm international guidelines on criteria for contraceptive use. However, it is expected that national programs will use the MEC to update or develop their own contraceptive eligibility guidelines in the light of their national health policies, needs, priorities and resources. The guidelines are reviewed and updated by the WHO every 4-5 years to incorporate new evidence. In the event that new evidence is generated in the interim that is inconsistent with the current recommendations, a technical consultation can occur to address the new findings. In light of the findings reported by Heffron *et al.* [1], which reported a ~2-fold increased risk of HIV acquisition among women using injectable contraceptives, the WHO scheduled a technical consultation to review all available evidence on this subject.

The WHO uses the classification system presented in Table 1 to make recommendations regarding the MEC for contraceptive use. These recommendations are informed by available evidence on worldwide hormonal contraceptive use, HIV epidemic patterns, biological mechanisms, systematic reviews and modeling studies. Available evidence is evaluated using the GRADE system for rating clinical evidence, which allows reviewers to assign a value to the evidence that reflects the level of confidence in the findings. Additional information regarding the details of the GRADE system, as well as strengths and limitations of this approach, can be found elsewhere [2, 3]. For the context of this meeting, one key component of the GRADE system was noted: this system begins by ranking data from observational studies as "low" relative to data from randomized trials; however, certain attributes of the body of evidence can raise the grade of observational evidence, and certain attributes can lower the grade of randomized evidence. At present, no randomized trial has been conducted to assess the effect of hormonal contraceptive methods on HIV acquisition. Our only evidence comes from observational studies, few of which were designed specifically to address this question. The majority of studies are secondary analyses of cohort study data or data collected as part of randomized trials for HIV prevention.

Table 1. WHO MEC Classifications

MEC Classification	Definition
1	A condition for which there is no restriction for the use of the contraceptive method.
2	A condition where the advantages of using the method generally outweigh the theoretical or proven risks.
3	A condition where the theoretical or proven risks usually outweigh the advantages of using the method.
4	A condition which presents an unacceptable health risk if the contraceptive method is used.
*	Added when number does not adequately capture the numerical guidance and more clarity needed (as recommended by an expert group).

The overall body of evidence for both injectable contraception and oral contraceptive pills and HIV acquisition was rated "low" (on a scale from high, moderate, low, to very low), due to

serious limitations and serious inconsistency between studies. With greater agreement among analytic approaches to this complex problem, it is possible that more consistent results would (organically) emerge, allowing for clearer interpretation of the data and enhancing the strength of the overall evidence base. The next WHO review of the evidence for the MEC will take place in 2014.

What is known about the competing risks of HIV and unintended pregnancy and maternal morbidity/mortality in various epidemiological contexts?

Dr. Phillips also presented on the results of three modeling studies assessing the competing risks of unintended pregnancy and maternal and child morbidity and mortality, potential for HIV acquisition and the impact of current contraceptive users discontinuing contraception among reproductive-aged women. In an analysis conducted by Rodriguez *et al*., the authors reported that discontinuation of injectable HC, without replacement with another method, or with replacement with combined oral contraceptive pills, would avert new cases of HIV but would lead to increased maternal mortality and decreased life expectancy [5]. If all women switched from injectable to IUDs, both new HIV cases and maternal mortality would be averted; however, 70-90% of all women would need to switch to prevent overall loss in life-years. Jain *et al*. reported similar results for maternal mortality and unintended pregnancy with the discontinuation of injectable HC in the absence of another highly effective methods of contraception [6]. Butler *et al.* reported that if there is an association between use of injectable HC and HIV acquisition, based on the current data, 27-130,000 infections per year could be attributed to injectable contraceptives, predominantly in southern/eastern Africa [7]. Similar to the findings described above, reducing injectable use globally could result in an increase in births and maternal mortality. The impact of reducing injectable use would depend on the true magnitude of an association with injectable contraception and HIV risk. A net public health benefit to removing injectables is unlikely, except in countries with large HIV epidemics, such as South Africa. In summary, each of the three modeling studies implemented different methods, but obtained similar results: discontinuation of injectable HC, without replacement with another highly effective method, would result in an increase in births in all settings and a loss of life-years except in certain settings that continue to have a high HIV incidence.

What is the evidence from observational studies regarding the risk of HIV acquisition among women using hormonal contraceptive methods?

Dr. Sandi McCoy and Ms. Lauren Ralph provided an overview of evidence from observational studies that assessed the effect of HC on HIV acquisition. Their overview also explored the heterogeneity in the evidence as it relates to the study population, study design features, analytic methods, and the effect of this heterogeneity on the validity of effect estimates. The review included 24 studies from 22 different study populations. There were temporal trends in the design of studies to assess this issue. Earlier studies published in the 1990's analyzed data from prospective cohort or nested case-control studies. These studies were small and contained minimal disaggregation of specific hormonal contraceptive methods. Starting around the 2000's, studies include larger sample sizes and more refinement in defining hormonal contraceptive exposure. All studies published in the last year (2012) that have assessed this issue have used data from HIV prevention trials.

The presenters identified 17 reports from 14 studies that assessed the effect of oral contraceptives on HIV acquisition risk. Three reports estimated a statistically significant increased risk in HIV acquisition with oral contraceptives (effect estimates = 1.46, 1.50 and 4.50), four reported a moderate though not statistically significant association (effect estimates 1.19-1.80), while all others report non-statistically significant effect estimates around or below 1. A total of 16 reports from 13 studies assessed effect of injectable HC and HIV acquisition: 5 reports assessed any injectable exposure (DMPA or NET-EN), 13 assessed the effect of DMPA, and 5 assessed the effect of NET-EN (total >16 since studies reported results for multiple exposures). Of the studies assessing the effect of DMPA, 7 studies reported a statistically significant increase in HIV risk among women who reported using DMPA (~1.5-3.8 fold). The remainder of the studies showed no statistically significant association and report effect estimates both above and below 1.0. The results from studies assessing the effect of NET-EN report no statistically significant associations, though there are fewer studies and heterogeneous effect estimates. A meta-analysis currently being conducted by the presenters includes results from 10 studies that assess the effect of DMPA on HIV acquisition.

Based on the data presented in the review, current evidence does not support an increase in HIV risk among women using oral contraceptive pills. The evidence is mixed for the risk of HIV acquisition among women using DMPA, but does not rule out the possibility of an effect on HIV acquisition risk. The evidence does not appear to support an increase in HIV risk among NET-EN users; however, additional data are needed. Data were not available for any other hormonal contraceptive methods; therefore, no conclusions can be made about the effect of other hormonal methods on HIV acquisition risk.

In addition, the presenters reviewed areas of heterogeneity in the body of evidence as summarized below. These differences between studies are important to consider as they impact the interpretation of the individual results.

- *Study population*: The majority of studies included populations in Southern and Eastern Africa. The underlying risk of being exposed to HIV is likely to have varied greatly as a result of the study populations included in these analyses. In addition, the type of population and potential for HIV exposure varied among the studies: 12 studies included data from women from the general population who were classified by the presenters as having minimal or heterogeneous risk of HIV exposure; 8 studies included data from sex workers which were considered at high risk for HIV exposure; and 2 studies included data from women in sero-discordant partnerships which were considered to have a guaranteed risk of HIV exposure. In discussions regarding the heterogeneity of the study population, the importance of the potential for HIV exposure was introduced. For example, for non-partner HIV prevention trials, there is increasing discussion and acknowledgement that not all women enrolled in the trial will be exposed to HIV. This heterogeneity in risk of HIV exposure and a woman's perception of risk, may impact decisions regarding choice of contraceptive methods and condom use.
- *Study design features*: There was substantial variation in the interval between study visits, which impacts the likelihood of misclassification of exposure, outcome, and other variables. Duration of follow-up also and rates of loss to follow-up also varied substantially between the studies. Of the studies conducted to date, only three (Martin 1998/Lavreys 2004/Baeten

2007, Kleinschmidt 2007, Morrison 2007/2010) were prospective studies designed specifically to assess the relationship between HC and HIV risk; 11 analyzed data from cohort or case-control studies, and 9 were secondary analyses of completed randomized trials. Of the known planned analyses, all will use data from completed randomized HIV prevention trials.

- *Analytic methods*: A range of analytic methods have been used to assess this relationship. Among studies that adjusted for potential confounders, no two studies adjusted for the same set of confounders. The presenters felt that age and condom use represented a minimally sufficient set of adjustment cofounders when considering which studies should be included in a meta-analysis; however, studies assessed condom use in a variety of ways. Condom use measures are prone to mis-measurement and misclassification to due recall bias and social desirability. There was discussion about the general reliability of condom use measure, since in some studies condom use is not associated with a decrease in HIV risk as would be expected. More recent studies have used marginal structural models (MSM) in addition to Cox proportional hazards models; however, there were also differences in the implementation of MSM. Lastly, the comparison group was not uniform between studies. The authors describe two commonly used comparison groups: type 1 compares participants using the HC method of interest compared to participants using no contraceptive method or non-hormonal contraceptive methods (i.e. participants reporting condoms, IUDs, or other non-hormonal methods); type 2 compares participants using the hormonal contraceptive method of interest to all other participants (i.e. oral contraception vs. all other hormonal and non-hormonal methods).

Definition of "total effects" and "direct effects"

These terms total and direct effects come from the causal inference literature and were used frequently throughout the meeting, especially as they relate to identifying the appropriate comparison group and applying and interpreting different statistical approaches. General definitions of these terms are provided below along with an example. Figure 1 displays a simplified causal diagram for the relationship between an exposure and outcome of interest. In this diagram, the direct effect is represented by "C", in that it is the effect of the exposure on the outcome that is not accounted for by the mediator variable. The total effect is represented by the direct effect "C" combined with the effect of the mediator "AB" (not necessarily a linear combination). Thus the total effect represents the direct effect of the exposure on the outcome and the effect of the exposure on the outcome through the mediator variable. In the context of this meeting, let the exposure = DMPA use and the outcome = HIV acquisition. For illustrative purposes, let the mediator = condom use, assuming condom use is lower in women using hormonal contraceptive methods including DMPA and lack of condom use could lead to an increased risk of HIV acquisition (in practice, condom use could operate as a mediator, confounder, or both). Based on Figure 1, the direct effect is the effect of DMPA use on HIV acquisition not mediated by condom use. The total effect is the overall effect of DMPA use on HIV acquisition, either through the hormones in DMPA or due to changes in condom use as a result of use of a highly effective contraceptive method.

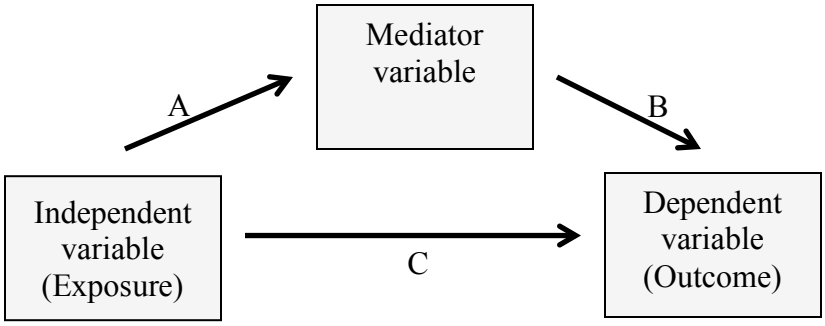

Figure 1. Conceptual model for total and direct effects

The terms total effect and direct effect denote what is being measured through statistical models; however, there was also a great deal of discussion and interest in understanding the "biologic effect" of HC on HIV acquisition. The biologic effect can be thought of as the effect of the exposure in the absence of any confounders or mediators. For example, using a primate model, comparing HIV acquisition risk among primates exposed to DMPA versus not-exposed following an HIV challenge. If a true "biologic effect" exists, the public health response might be to more strongly emphasize expansion of women's contraceptive choices so they could choose an alternative highly effective method if they are at high risk for HIV. Alternately, if the association is driven by changes in sexual behavior (such as reductions in condom use), the response might instead focus on improving condom use in conjunction with highly effective contraceptive methods. Due to the presence of confounding and mediators in studies in humans as well as ethical issues in study design, it is likely impossible to fully isolate the biologic effect. However, it may be possible to approximate the biologic effect either by conducting a randomized trial (not discussed here) or by estimating the direct effect of HC on HIV risk in observational studies (discussed in detail below). Moving forward in the meeting, the focus centered around research questions and analytic methods that compare the total effect and the direct effect.

What is the comparison group of greatest interest?

Dr. Chelsea Polis and Dr. Jared Baeten debated the scientific and policy implications of different comparison groups to assess the relationship between HC and HIV risk. Different information regarding the effect of HC on HIV acquisition risk can be generated depending on the composition of the comparison group. In addition, various analytic approaches may provide different information concerning the total versus direct effect of HC on HIV risk. Based on the presentation and accompanying discussion it became clear that there are a number of different questions that can be asked within the broader context of assessing the relationship between HC HIV acquisition. Though individuals may prefer the information generated from some questions over others, it is important that researchers investigating this topic carefully consider and communicate what effect estimate they seek to obtain in their analysis in order to improve the evidence base and inform policy recommendation.

"Non-hormonal methods" comparison group

On the one hand, it can be argued that in order to assess the effect of HC on risk of HIV acquisition, it is necessary to compare users of hormonal contraceptive "A" to women not using HC methods "A, B, or C", since the baseline risks of HC methods "B" and "C" are unknown. In past studies, the composition of the non-hormonal comparison group has varied, with some women using a highly effective non-hormonal method (e.g., sterilization), others using a less

effective non-hormonal method (e.g., condoms or withdrawal), and others not contracepting at all. In some of the first studies to assess HC and HIV risk, women using alternative HC methods were included in the comparison group. Some of the non-hormonal methods, such as condoms, prevent HIV, complicating analyses since women using HC are typically less likely to use condoms, and also less likely to use them consistently. Studies suggest that women who use condoms for contraception use them more consistently than women who use condoms primarily for HIV/STI prevention. It is important to understand the composition of a non-hormonal comparison group as it relates to confounding. For instance, women not contracepting may have very different behaviors from contracepting women. In many studies, the composition of comparison groups has been a mixture of non-hormonal method users and non-contraceptors.

The use of a non-hormonal comparison group may be beneficial for both analytic and practical reasons. If confounding can be adequately addressed, which is admittedly a challenge in observational epidemiology, using a non-hormonal comparison group should theoretically allow an isolation of the biological effect of HC on HIV susceptibility. In addition, there is very limited evidence regarding the impact, positive or negative, of other contraceptive methods. For example, comparing DMPA users against sterilized women could theoretically provide a straightforward interpretation of the effect of HC, whereas comparing DMPA users only against women using implants may be less informative. If both DMPA and implants similarly increase risk of HIV acquisition, the risk ratio could appear null. That said, depending on the composition of the comparison group, there may be some tradeoffs between the potential for confounding and isolation of any effect of the hormonal contraceptive method. However, this comparison is also important given the current global realities regarding access to hormonal contraceptive methods. In many settings, there are limited hormonal contraceptive options; therefore, a comparison between two hormonal contraceptive options is less informative in a setting in which one of those methods is not available. For example, injectables represent a large proportion of method mix in sub-Saharan Africa. This predominance of injectable HC in this region is likely a combination of preferences for injectables and issues of access, cost, service delivery, and provider bias against other methods. Even for women with access to more options, personal preferences about various methods may limit those she views as acceptable to her. Analyses using a non-hormonal comparison group can provide cross-method analyses. For example, one study can look at the effect of hormonal method "X" versus a non-hormonal comparison group, and also at the effect of hormonal method Y versus a non-hormonal comparison group, and then compare the risk estimates for "X" and "Y" to provide cross-method comparisons to women and providers. Finally, understanding the effects of various hormonal contraceptive methods as compared to women not using HC is important not only for developing contraceptive counseling and guidelines, but also for the development of future contraceptive and multipurpose prevention technology products.

Alternative comparison groups besides the "non-hormonal methods" group

On the other hand, it can be argued that the non-hormonal comparison group may not be ideal, since women who choose to use HC may not be the same as those who chose to use non-hormonal methods or no method. The argument for the use of a non-hormonal comparison group assumes that it is possible to fully address confounding when comparing women using HC to women who have chosen not to use HC. The women in a non-hormonal comparison group may represent a mix of women using other highly effective contraceptive methods, less effective

contraceptive methods or no methods. Arguably, these subgroups are different from each other and potentially quite different from hormonal contraceptive users in terms many important factors that are difficult to measure, including fertility plans, social stability, HIV exposure, not to mention behaviors related to sexual activity.

Comparisons across hormonal contraceptive methods can be informative for women, providers and policy makers. For women with access to a variety of contraceptive methods who chooses to contracept, her choice may not be method "X" versus nothing, but rather method "X" versus method "Y" versus method "Z". Both the woman and the provider require information regarding cross method comparisons in order to weigh the risks and benefits of "X, Y and Z". Some meeting participants argued that analyses with a non-hormonal group feed the argument that the alternative to a hormonal contraceptive that potentially increases the risk of HIV is no HC at all. For policy makers, cross-method comparisons can help to inform decision-making regarding which methods should be prioritized.

It can be argued that the ability to choose between multiple hormonal contraceptive methods is not a reality for many women living in regions highly affected by HIV. That is certainly true in many settings and all meeting participants agree that the contraceptive mix needs to be expanded. However, in some settings, such as South Africa where the HIV incidence among reproductive aged women continues to be high and the strongest association between DMPA and HIV risk has been observed, alternative hormonal contraceptive methods, such as NET-EN, do exist. Comparisons between these two readily available methods in this population could directly impact policy in South Africa and, depending on the findings of such a comparison, could influence policy in other countries as well.

What is the difference between the "total effect" and the "direct effect"? How well can we estimate either and which one are we most interested in?

Clarifying the extent to which effects observed in prior observational studies are direct or confounded/mediated by measured behavioral factors and specifying which effect will be assessed in future analyses will help the field better interpret the evidence generated from HC-HIV analyses using observational data.

Isolating the direct effect of HC

There is a clear need to try to isolate the biological effect of HC on HIV susceptibility, through methods that estimate the direct effects of this relationship. For example, consider a woman with limited access to contraceptive options, whose partner refuses to wear a condom. She wants to know if using DMPA increases her risk of HIV acquisition. If we fail to separate the direct from total effects, all we can say is that using DMPA may increase her risk of HIV, but we don't know if it is because of the hormones in a DMPA, or because DMPA users (on a population level) are less likely to use condoms. Since she is already unable to negotiate condom, she may now be left with no acceptable, available, and appropriate contraceptive option. Furthermore, since condom use is likely to be inversely associated with any highly effective contraceptive method, failure to isolate biological effects may tell us little about the relative impact of various highly effective contraceptive methods.

Assessing the total effect of HC

In assessing the general question of dose hormonal contraceptive use increase HIV risk, what we are trying to measure is a complex mix of biology (e.g., physiological effects of HC) and behavior (factors that may confound and/or mediate the association between HC and HIV risk). For women, providers and policy makers, in the end, it is an outcome (HIV acquisition) that is important, as well as the total health of the woman (including events such as unintended pregnancy). It is unclear if observational epidemiology will ever be able to estimate the biologic effect. If the biologic effect could be isolated, it may represent a "false purity", since it does not capture the totality of risk actually faced by women.

Conceptual model for factors that may influence the relationship between HC and HIV risk

Directed acyclic diagrams (DAGs) specify hypothesized causal, confounding and mediating factors that may influence the relationship between an exposure and outcome of interest [8]. Meeting participants, led by Dr. Daniel Westreich, drafted a DAG for the relationship between HC (exposure) and HIV acquisition (outcome). While it was not feasible to specify one consensus DAG given the large number of potential variables of interest, there was general agreement that key factors exist which should be carefully considered as potential confounders, mediators, or both. Furthermore, there was general agreement that several variables could act as time-varying confounders. Here, we present a simplified conceptual model that includes a list important potential confounders and mediators (Figure 2). Future analyses should consider assessing these factors.

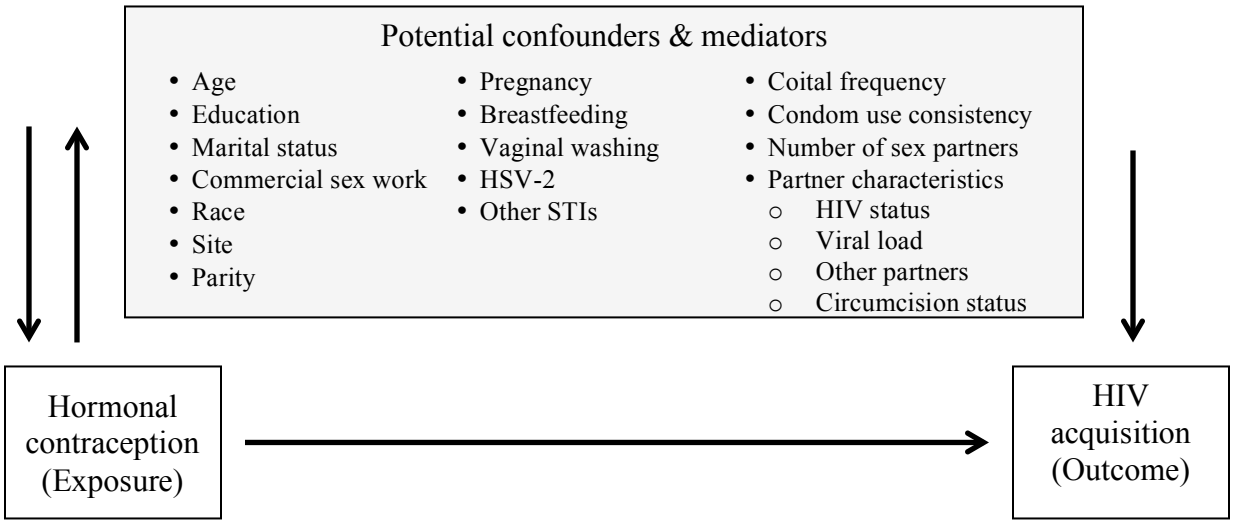

Figure 2. Factors that may confound or mediate the relationship between hormonal contraceptive use and HIV acquisition

Causal inference approaches to HC-HIV analyses

Dr. Erica Moodie presented on causal inference approaches to assess the relationship between HC and HIV acquisition. The majority of prospective studies have assessed the effect of HC on HIV risk using Cox proportional hazards models. These models provide a measure of association of the total effect adjusted for potential measured confounders. One limitation of this approach is that these models can't adequately address issues of time-varying confounding and mediation. Time-varying confounding and/or mediation can occur when covariates at time *t* predict future outcomes/covariates as well as subsequent treatment, and are themselves affected by prior treatment. Using the DAG presented in Figure 3 as an example, assume A=DMPA use, L=condom use and Y=HIV acquisition. Condom use at L_1 is a potential confounder since it is associated with A_2 and Y_2.

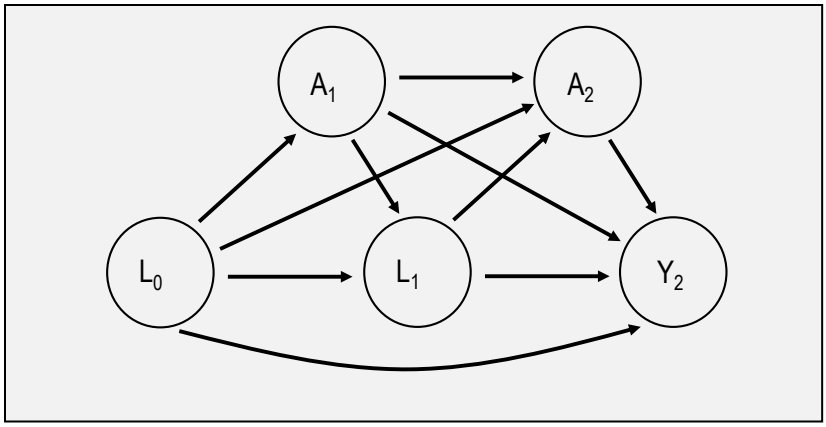

Figure 3. Example DAG with time-varying confounding (presented by E Moodie)
A= DMPA (exposure) Y = HIV acquisition (outcome) L = Condom use (confounder)

Condom use (L_1) is a collider since it is a mediator for the prior condom use (L_0) and DMPA use (A_1) pathways. Conditioning on the collider (L_1) will result in what is called collider stratification bias, potentially generating a spurious association between the exposure and the outcome. Alternative approaches are required in order to address potential confounding by condom use at L_1 without introducing bias through conditioning on a collider.

Assessing the total effect through traditional marginal structural models

Marginal structural models (MSMs) can be used to give an estimate of the total effect, adjusted for time-varying confounders, that is less biased since it addresses time-varying confounding [9-11]. These models rely on counterfactuals: outcomes that would have been observed had a participant been exposed to a particular treatment pattern. An MSM is a model for the average outcome if the entire population was exposed to a particular treatment pattern, for each possible treatment pattern under study. Since only one of the X possible counterfactuals can ever be observed, this can be re-framed as a missing data problem, whereby inverse probability weighting can be used to up and down weight individuals to create a pseudo-population of subjects whereby treatment receipt is not affected by time-varying covariates (no time-varying confounding). Because there is no confounding, there is no need to condition on L_1 (from the example above) and by not conditioning on L_1, we do not block mediated pathways or induce collider-stratification bias.

A three-step process can be used to estimate a MSM [12]. Details regarding the construction and implementation of MSMs with longitudinal data have been presented [8, 10]; however, a brief summary of the process is provided below. First, a treatment model is fitted, such as a logistic regression model, for the probability of being treated at each interval. Next, the model is used to predict the probability that a person received the exposure pattern that they did in fact receive by

taking the product of the probability of receiving the observed treatment in each interval. An inverse probability weight for each individual is generated by setting each individual's weight to one over the probability that a person received the exposure pattern that they did in fact receive. It is recommended that the weights be stabilized, normalized or truncated in order to produce "well-behaved" weights (mean =1, small range) that will result in a small variance of the effect estimate [13]. Lastly, standard software can be used to fit a regression model for the outcome given the exposure weighting each individual by the inverse probability weights. The resulting effect estimate is unbiased, provided a number of assumptions have been met, detailed below.

It is important to note the assumptions required when implementing MSM: (1) there must be no unmeasured confounding at all intervals; (2) the treatment model and the response model need to be correctly specified; (3) the exposure status is not uniquely defined by covariates (called the positivity or experimental treatment assignment assumption) since even non-zero but very small treatment probabilities can cause serious instability in estimates; (4) the exposure must be well-defined; (5) the participant outcomes are independent (i.e. no interference between participants, so that each person's potential outcome depends only on their exposure, but not that of others). Note that these assumptions – with the exception of correct specification of the treatment model – must also be met to obtain causally-interpretable estimates from standard regression models, which may be used in settings where exposures are not subject to time-dependent confounding.

Assessing the direct or indirect effect through mediation analyses

Standard implementation of MSM addresses bias due to time-varying confounders and should produce an unbiased estimate of the total effect. A different implementation of MSM can be used to understand mechanisms of an exposure by decomposing the exposure into direct and indirect effects. As defined previously, direct effects represent the part of the exposure effect that does not act through a given set of potential mediators ("C" in Figure 1). The indirect effects represent the part of the exposure effect that acts through that given set of potential mediators ("AB" in Figure 1). In mediation analyses, it is important to clarify the type of "direct effect" being assessed. Two frequently used types are controlled direct effect (CDE) and natural direct effects (NDE). A controlled direct effect is the average effect of exposure on the outcome if the mediator were controlled (set at a specific value) uniformly across the entire population. For example, with partners' infectivity and condom use as mediators, what is the effect of DMPA if all partners have the same viral load and use condom use 80% of the time? The natural direct effect is the average effect of exposure on the outcome if the mediator were set to the value it would have taken in the absence of exposure. For example, the effect of DMPA if partners' viral load is the same in the absence of DMPA and condom use is the same in the absence of DMPA. In the NDE, the value of the mediator is set to a counterfactual outcome, which allows 'natural' variation in the level of the mediator across subjects, but relies on never-observable counterfactuals.

How a mediation analysis is performed depends on what type of effect is sought (controlled vs. natural), and what assumptions one is willing to make. All mediation analyses require that the following assumptions are met: (1) there is no unmeasured confounding between the exposure and outcome; (2) there is no unmeasured confounding between the mediator and outcome; (3) no confounders of the mediator and outcome are caused by the exposure; (4) exposures must be well-defined; (5) the participant outcomes are independent; and (6) the model is specified correctly. The above assumptions are sufficient to identify controlled direct effects using

standard regression. To identify natural direct effects, there must also be no interaction between the mediator and the exposure.

In summary, MSMs are straightforward to compute and allow estimation of a range of parameters including population average effects under specific exposure patterns (total effects), and the decomposition of effects into the effect mediated through a variable (direct effects) and the "remainder" of the effect that is not mediated through the primary variable of interest (indirect effects).

Assessing the direct or indirect effect through g-methods

In addition to marginal structural models, other advanced methods such as generalized methods (g-methods), can be used to estimate causal effects in the presence of time-varying confounders that may be affected by treatment. Dr. Daniel Westreich presented an overview of g-methods and how they can be applied to HC-HIV analyses. There are 3 main g-methods: parametric g-formula, nonparametric g-formula and g-estimation with structural nested models (semi-parametric). Non-parametric g-formula can break down in realistic data and g-estimation with structured nested models requires linear outcomes and is difficult to use in mediation analyses; therefore these methods may not be appropriate for HC-HIV analyses. However, the parametric g-formula could be used to complement the results of MSM. What is modeled in in MSM is the exact opposite of what is modeled in parametric formula, thus use of both models can help triangulate effect estimates.

Additional thoughts on analytic approaches to HC-HIV analyses

Dr. Stephen Cole presented on and moderated the accompanying discussion on analytic epidemiology as it relates to assessing the relationship between use of hormonal contraceptives and HIV risk. Dr. Cole outlined a series of steps that are recommended when designing and conducting analyses. A complete list of the steps is presented in Box 1; however, the presentation and discussion focused primarily on steps 1-3 (as summarized below), as these steps are the most critical for the analytic design and serve as the foundation for all additional steps. The steps are intended for use in secondary analyses of observational data, however, it was suggested that one approach the analysis as if they were designing a randomized trial. For example, in determining the study population or "context" (step 1), it is helpful to clearly define the eligibility criteria for being included in the analysis. Investigators are encouraged to clearly state the criteria for inclusion/exclusion in observational analyses.

Box 1. Steps in analytic epidemiology

1. Determine the context
2. Determine the exposure(s) of interest, ensuring well-defined, *
3. Select the outcome(s), *
4. Select the parameter(s) of interest, *
5. Determine an accurate estimator(s) of that parameter, *
6. Ensure conditional exchangeability, *
7. Assess positivity, *
8. Ensure negligible measurement bias, *
9. Ensure correct models specification, *
10. Explore transportability/evidence-base, *

** Return to earlier step(s) as needed*

The next step includes determining the exposure(s) of interest, ensuring that it is well-defined. In terms of the primary exposure of interests, investigators must not only define what it means to be exposed but what it means to be not exposed. As discussed above, the choice of comparison group has implications for what questions can be assessed. Investigators must also clearly define

other important exposures that may confound or mediate the relationship between the primary exposure and the primary outcome. Finally, the outcome of interest is selected. In secondary analyses, it is important to determine if there are enough outcomes in order to adequately assess the question of interest. In addition, it is important to consider the timing/frequency of outcome measurement and other potential competing risks for the outcome, such as pregnancy (discussed below). By using a systematic approach to secondary analyses using observational data, such as the approach described in Box 1, we can feel confident that the analytic epidemiologic methods are sound. However, inherent limitations in the data that result in unmet assumptions may result in bias or error. It is important for investigators to be aware of departures from these assumptions and conduct sensitivity analyses accordingly.

Analytic treatment of pregnancy

The issue of pregnancy in HC-HIV studies is a particular challenge since it is common and has also been associated with an increased risk of HIV acquisition in some, but not all, studies. In addition, it can be viewed as competing risk of sorts for women using HC in that it is an event that prevents the event of interest (HIV acquisition) from occurring under the original exposure conditions (i.e. women who are pregnant are no longer using HC). Studies have addressed pregnancy several ways: no adjustment for incident pregnancies, censoring at first positive pregnancy test, adjustment for pregnancy as a time-varying confounder in standard regression models, and weighting in marginal structural models. The most appropriate method to address pregnancy may depend on the question being asked. Dr. Angela Crook and Dr. Pai-Lien Chen presented a proposed approach to addressing pregnancy in the analysis of HC and HIV risk among women enrolled in the MDP 301 microbicide trial. The authors used a method for dynamic treatments proposed by Hernan *et al*, which builds on marginal structural models and uses "artificial censoring" – where women are censored at the point of deviation from their original treatment [14]. In this case, incident pregnancy represents a deviation from the original treatment of HC use. Inverse probability weights are then used to account for the probability of this censoring, and the outcomes are compared across uncensored individuals. The implementation of this method represents a novel approach to addressing pregnancy in HC-HIV studies. The results presented were preliminary and it is uncertain how the results generated using this approach compare to those generated using previously employed approaches. At present, the group could not come to consensus regarding the best analytic approach to address pregnancy in these analyses and recommend further study in this area. Sensitivity analyses are recommended with the implementation of different approaches to address pregnancy in order to gauge the robustness of the results.

Lessons learned from operationalizing HC-HIV analyses and implementing MSM

Dr. Charlie Morrison and Dr. Pai-Lien Chen shared their experience in conducting the analysis and re-analysis of a prospective cohort study designed to assess the effect of HC on HIV risk [15, 16]. One of the primary challenges in assessing this question is accurately ascertaining both the exposure and potential cofounders and mediators in relation to HIV exposure and seroconversion. This challenge is clearly illustrated in Figure 3 (developed by Jared Baeten). Some women are likely to use HC methods consistently while others may use these methods intermittently and their use may or may not correlate with periods when they are potentially exposed to HIV. There are few studies of long acting methods such as implants or hormone-containing IUDs. Studies evaluating these longer-acting methods may be less subject to

misclassification. When evaluating HC as an exposure, regardless of the specific method, the presenters argued that monthly visits allow for more precise measurement of the exposure and potential confounders.

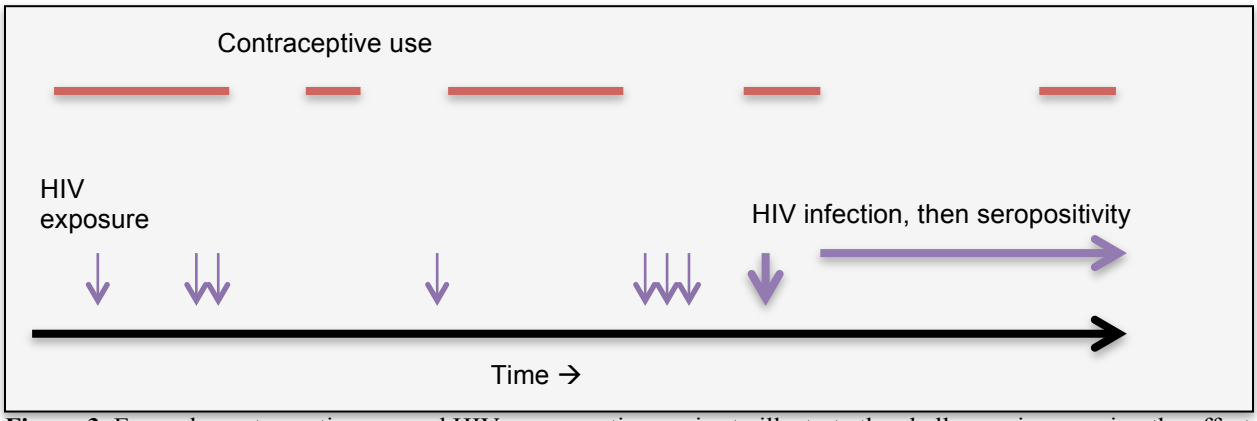

Figure 3. Example contraceptive use and HIV exposure time series to illustrate the challenges in assessing the effect of HC and HIV acquisition risk (developed by Jared Baeten)

Since the HC-HIV study was designed specifically to assess this relationship, data on adherence to HC was collected using a detailed contraceptive calendar. The presenters reported that, in their experience, there was more consistent hormonal contraceptive use when it was provided as part of the study at the study site. In addition, study staff can record what was dispensed instead of relying on participant self-report (i.e. DMPA injection given versus participant will get DMPA injection at the clinic at end of the week).

The presenters also discussed the details of their re-analysis of the data using marginal structural models and other sensitivity analyses [16, 17]. The authors chose to re-analyze the data using marginal structural models in order to address confounding by time-varying factors. These factors included condom use, characteristics of the primary partner (known HIV+, penile discharge, significant weight loss, sex with commercial partners, spent nights away from home during the past 3 months) and participant behaviors (multiple partners, new sex partner, commercial sex). The results from the MSM analysis and subsequent sensitivity analyses have been presented previously, thus will not be reviewed here [16, 17]. In the implementation of MSM, the presenters provided some additional thoughts on the construction of the weights used in MSM. In particular, they recommended truncating extreme weights in order to reduce bias-variance.

Dr. Sandy McCoy and Dr. Renee Heffron presented their experiences implementing MSM [1, 18], with a particular focus on key decision points in the implementation process. Dr. McCoy began with a discussion assessing potential time-varying covariates. Despite theoretical plausibility for factors such as condom use at last sex, circumcision status of the regular partner, and having a new sexual partner to be considered time-varying confounders, using a data driven approach, these factors were not associated with the exposure or the outcome in Dr. McCoy's dataset. In observational epidemiology, adjustment for potential confounders can be based on *a priori* assumptions as well as biologically plausible, data-driven relationship. In future analyses using MSM, it is recommended that the decision making behind how and why factors were considered time-varying confounders should be clearly stated. This is particularly important for

constructing weight models since the inclusion a large number of confounders may not improve the model. Including many unnecessary confounders may result violations of the positivity assumption for MSM, due to small or no observations in strata. In addition, it may introduce selection bias [13].

Dr. McCoy also discussed the challenge of time-ordering assumptions when performing a secondary analysis to assess the HC-HIV relationship using MSM. To calculate weights for MSM, the probability of using HC at a given visit must be predicted. To predict this probability, one must ensure the correct temporal order between possible confounders and HC use at a given visit. In general, covariate information through visit $t-1$, rather than through visit t, is used to predict hormonal contraceptive use reported at visit t. However, in secondary analyses, questions that seek to ascertain hormonal contraceptive exposure may not have been asked in a manner that provides precise exposure information for this type of secondary analysis. For example, in the question "What are you currently doing to prevent pregnancy?" it is unclear if this is capturing hormonal contraceptive use prior to the visit (backward in time) or at the visit and beyond (forward in time). There was a fair amount of discussion surrounding this point. The group recommended that for future trials or cohort studies where the investigators plan to assess the effect of HC on HIV risk, thought should go into how questions are asked regarding use of HC and potential confounders in order to optimize data collection for assessment of this question in the context of the larger study.

Missing data also plays a critical role in how analytic decisions are made. More recent studies assessing the effect of HC on HIV have used data from HIV prevention trials. Despite efforts to retain participants in these trials, missed visits (including multiple missed visits in a row) are common. For women with missing exposure histories it is not common to ask about exposures, covariates, and outcomes occurring during the missed visit interval. However, when using MSM, it is necessary to compute a weight that stretches across the missing time. In accompanying discussion about this issue, it became clear that how to address missing data when assessing the effect of HC on HIV risk, with standard Cox models or MSM, is an important topic for further discussion and study.

Dr. Heffron focused on the impact of different parameterizations of factors included in the weight model for MSM. Having *a priori* ideas about which key factors are important for inclusion in the weight model may guide the analysis. An exploration of different variable parameterizations and the inclusion of additional potential time dependent confounders can give the analyst a good idea of model robustness. In addition, Dr. Heffron supported the truncation of weights in order to balancing the advantages of having a small range of weight values with a mean close to 1.0 with the best control for confounding (as mentioned above and discussed in the literature[13]). Both presenters concluded with recommendations for how to report analytic decisions (Box 2).

Box 2. Recommendations to for how to report analytic decisions when assessing the effect of hormonal contraception on HIV risk using observational data

For all statistical models:

- Define factors assessed as potential time dependent confounders
- Describe how missing values were addressed
- Describe how time order was assigned

For marginal structural models

- Describe the parameterization of each variable included in the weight model – including details about prior exposure
- Give basic characteristics of weights (mean, standard deviation, minimum/maximum if space allows)
- Describe any techniques used to balance weight behavior and bias control (i.e. truncation)
- Describe the variables included in the final weighted model
- Creativity can help to fit within journal word count limits (use of table footnotes or a web appendix)

The presentation of lessons learned from the implementation of MSM lead to a broader discussion about the use of Cox and MSM in HC-HIV analyses. With the exception of the observational HC-HIV study [15, 16], studies that have implemented Cox models as well as MSM have reported little difference between the models. If Cox models suffer from collider stratification bias as a result of time-varying confounders that also operate as mediators, MSM should provide a different, less biased, result. One potential reason for the similarities in the results is that both models are removing time-varying confounding equally by adjusting (Cox model) as by weighting (MSM), thus the lack of difference. In addition, if there is no strong mediator (or if the strong mediator is not captured accurately), then the results will also be similar. Alternatively, it is possible that both models (Cox and MSM) are being either specified or mis-specified correctly. Lastly, there may be minimal difference when assessing immediate or short-term exposures. In many studies, the exposure assessed occurs in the interval immediately preceding the outcome (i.e. DMPA use report at the quarterly visit prior to detection of HIV seroconversion). Studies that assess the cumulative exposure or long term exposure may, in theory, observe larger differences between the models.

Missing data, inaccuracies, and suboptimal wording of questions present significant challenges for HC-HIV analyses. When preparing data for analysis, key decisions are required by the team – the consequences of which are not always black and white. In addition, it is important to acknowledge that MSMs have important assumptions that may not be met, depending on the dataset. Effectively communicating analytic decisions will help improve our understanding of the comparability and interpretation of future HC-HIV risk analyses.

Discussion of ongoing and upcoming HC-HIV analyses

Data from two completed analyses and information from two planned analyses were presented to the group. The completed analyses are not yet published; therefore, exact effect estimates are

not presented in this report. However, a summary of the presentations, general findings and accompanying discussion is presented below.

HPTN 035

Dr. Jennifer Balkus and Dr. Elizabeth Brown presented results from an analysis that assessed the effect of HC on HIV acquisition among women participating in HIV Prevention Trials Network Protocol (HPTN) 035, a phase II/IIB microbicide trial that evaluated the safety and efficacy of BufferGel and PRO 2000 gel. The analysis included 2804 participants from Malawi, South Africa, Zambia and Zimbabwe. Cox proportional hazards models were conducted to assess the association between HC and HIV risk. Overall, compared to women not using HC, the presenters reported a marginal, but non-statistically significant increased risk in HIV acquisition among women using injectable hormonal contraception (DMPA or NET-EN) and no increased risk in HIV acquisition among women using oral contraceptives. Using data from 4 countries and 7 different sites, the presenters also discussed the heterogeneity of the data with regard to HC use, potential confounders and HIV incidence. To illustrate the effect of this heterogeneity, effect estimates were also presented by country, though due to few events at some sites, the confidence intervals around the effect estimates were extremely wide and effect estimates should be interpreted with caution. Adjustment or stratification by site or country typically serves as a proxy for a number of other unmeasured or unknown factors relevant to the population enrolled in the study at that site. However, based on the data presented stratification by site/country may not be capturing the full effect. The presenters wanted to call attention to this issue in the interpretation of the results from their analysis and for others as they embark on future analyses using data that may contain similar issues concerning heterogeneity. In addition, the presenters felt that due to unmeasured confounding, MSM were not appropriate for use with this data. Preliminary MSM models produced identical results to those generated with adjusted Cox models.

MDP 301

Dr. Angela Crook summarized the results of MDP 301 trial, a phase III that evaluated the efficacy of PRO 2000 gel for HIV prevention. This analysis included 8,663 women from 6 sites in 4 countries (South Africa, Tanzania, Uganda and Zambia). Participants were asked about contraception and sexual behaviors at monthly visits and HIV testing was performed at 12, 24, 40 and 52 weeks of follow-up. Cox proportional hazards models were used to compare the risk of HIV between HC (DMPA, NET-EN, and oral contraception, each analyzed separately) versus women using non-hormonal methods. These results have not been published and will be presented at CROI 2013. The presenter reported that there was also a great deal of heterogeneity by site with regard to HC and potential confounders (similar issue to that reported in the HPTN 035 analysis); however, the effect estimates were similar across sites (data not presented). Additional analytic work is ongoing as there are plans to fit marginal structural models and conduct sensitivity analyses that explore other approaches to handling pregnancy.

VOICE

Ms. Lisa Noguchi presented plans for an analysis of data from VOICE, a five-arm phase IIB trial of two antiretroviral-based HIV prevention approaches in women (oral tenofovir, oral Truvada, vaginal tenofovir gel, oral placebo, vaginal placebo gel). 5,029 women from Uganda, South Africa and Zimbabwe participated in the trial with follow-up completed in late 2012. The results

of the primary analysis will be presented at CROI 2013; therefore, no information beyond baseline data was presented at this meeting. HC use, sexual behaviors and HIV status were assessed monthly. Contraceptive logs were used to clearly track contraceptive dispensing and switching between methods. Data from the logs can be used to differentiate between DMPA and NET-EN use in South Africa. In addition, women were required to use a highly effective method of contraception, such as injectables, oral contraceptives, implants or sterilization (IUDs not widely available at the study sites), in order to participate in the trial. As a result, the proportion of women who are not using HC is very small and will not be an appropriate comparison group. Ms. Noguchi is considering focusing the analysis on a comparison of HIV risk between women using DMPA versus women using NET-EN. The analysis would be restricted to participants from South Africa since only those participants had access to both injectable methods. The majority of participants enrolled in VOICE were from South Africa (N=4077) and, based on epidemiologic data from South Africa, it is likely that South Africa will contribute the most HIV endpoints to the analysis; therefore, an analysis restricted to South African participants should be sufficiently powered (minimum detectable effect size not known since results from the parent trial are not public).

Though this study can't assess the "direct effect" of either injectable method on HIV risk, it provides critical information on the risk of DMPA relative to NET-EN. For example, if the risk HIV is greater among DMPA users compared to NET-EN users, policy makers may choose to recommend that women at risk for HIV who desire HC be prescribed NET-EN, rather than DMPA. Therefore, information that informs the risk of one HC method relative to another may be important for policy, clinical practice and individual decision-making. That said, in discussions about the proposed analysis, concerns were raised about comparing women using injectable methods to women using oral contraception. Use of a highly effective contraceptive method was required for participation in the trial and data were not collected at screening concerning women who initiated HC use in order to be eligible for the trial. Women who don't really want to be on an effective form of contraception may choose oral contraception, since they can exert control of adherence (i.e. dispensed but not taken). Since data are not collected on adherence to oral contraceptives, there is the potential for misclassification in the oral contraceptive group. For example, if there were an equivalent risk of HIV among DMPA users and oral contraceptives users, due to differential misclassification among oral contraceptive users, we could incorrectly observe an increased risk of HIV among DMPA users compared to oral contraceptive users. Since adherence after dispensing is not an issue for women using injectable methods of contraception, the presenter plans to focus the primary analysis on the DMPA versus NET-EN comparison.

Individual participant data meta-analysis

Dr. Charlie Morrison and Dr. Pai-Lien Chen presented on their plan to conduct an individual participant data meta-analysis to evaluate the effect of HC use on the risk of HIV acquisition. The presenters plan to separately assess the effect of DMPA, NET-EN, and combined oral contraceptive pills compared to a non-HC group. Additional analyses will be conducted to assess effect modification by age and by HSV-2 status at baseline. Motivations for the individual participant data meta-analysis include increased statistical power to examine the HC-HIV risk relationship between important sub-groups (young women, HSV-2 negative women), investigation of confounding by both participant and study-level characteristics that may aid in

understanding the heterogeneity of study results, the potential for reduced bias due to differences in analytic strategies used by individual studies, and exploration of intra/inter-study variation in HC effects on HIV risk. The presenters plan to include studies that measured HIV prospectively at multiple time points with a testing interval of 6 months or less, had at least 15 incident HIV infections, measured HC use prospectively at multiple time points with a measurement interval of 6 months or less, and measured age, condom use and number of sex partners (minimum set of potential confounders). Studies with a significant amount of HIV and HC missing data (5%) and small sample size will be excluded from the analysis. The presenters plan to censor individuals if they switch contraceptive methods to use a non-study exposure (IUD, POP, or implant) and, in the primary analysis, they will censor at pregnancy. The presenters reported that the meta-analysis will include data from 18 studies with 36,500 HIV-seronegative women and over 1,800 HIV endpoints. The data from individual studies have been collected and prepared and analysis be commence in 2013 with results expected in late 2013.

A number of sensitivity analyses are planned with regard to switching of methods, different approaches to handling pregnancy, stratification by measurement of the exposure (i.e. by the ability to distinguish between types of oral contraceptives and types of injectables), and stratification by HIV incidence in the study population (high versus low incidence populations). The authors are also considering an analysis among the subset of women reporting no condom use. This potential analysis sparked a good deal of discussion. There is interest in looking at this subset of women due to concerns over the reliability of condom use data. Due to recall bias and social desirability, it is thought that women are more likely to over report condom use rather than underreport. Therefore, an analysis restricted to women who report no condom use may to less likely to suffer bias due to misclassification. If this analysis is conducted, the group suggested that the determination of no condom use be determined at baseline and then individuals become censored when they begin to use condoms. It is inappropriate to look forward into the data to see if women used condoms over the course of the study. In addition, there are also concerns about the composition of women who report no condom use, especially since condom use is frequently greater among women who do not use HC.

Adjunct analyses and additional data collection to support primary HC-HIV analyses

Dr. Renee Heffron and Dr. Jared Baeten presented on additional steps that could be taken to improve HC-HIV analyses and approaches that may contribute to overall understanding of the relationship between HC use and HIV risk. In an ideal setting, studies assessing this issue would be able to perfectly capture contraceptive use, obtain fully accurate characterizations of confounding factors, particularly sexual behavior, capture of all potential confounding factors and observe a sufficient number of HIV-1 seroconversions, including by different contraceptive types and within subgroups, so that study power is not limiting. However, the reality is that there is the potential for inaccuracy of self-reported HC use, self-reported sexual behavior (coital frequency and condom use) and potentially different amounts of inaccuracy among women using and not using HC. In addition, in most studies, there is an unknown level of actual exposure to HIV, difference across study sites and, too often, limited statistical power to detect associations among sub-groups. The presenters suggest potential improvements in 4 areas: (1) improved measurement of contraceptive use; (2) validation of sexual behavior reporting; (3) strategies to understand HIV-1 exposure; and (4) analyses to explore by-site effects of HC on HIV risk.

- *Improved measurement of contraceptive use.* Most studies use self-report of HC use, which is prone to misclassification through misreporting by the participant (unintentional or intentional), mis-timing (active switching between contraceptive methods or passive switching between using and not using the same method), and misleading (social desirability to be on or not on contraception or condoms). As an operational strategy to improve measurement of contraceptive use, study sites should provide contraception on site. For oral contraceptive, adherence data could be collected, including contraceptive diaries or returning pill packs. An analytic strategy to better understand adherence could include an analysis of pregnancy rates as a proxy for contraceptive adherence. Pregnancy rates that are in line with expected "real-world" rates for the contraceptive methods could be a marker of adherence. Some studies have reported similar pregnancy rates between women reporting oral contraceptive use and non-hormonal methods, which could be an indication of poor adherence/non-use of oral contraception. Lastly, measurement of exogenous hormones in all participants or a sub-set of participants could be implemented as a biologic strategy to gauge the degree of accurate reporting.

  There were several discussions regarding improved measurement of contraceptive use. A number of meeting participants were in support of having contraception dispensed at the study site, as it is better to have more control over dispensing rather than working on adherence, since measures of oral contraceptive adherence are traditionally poor. In addition, there was a discussion around how adherence to the contraceptive method has an implication for which type of effect estimate should be obtained. Though some participants may not be fully adherent to taking oral contraception, some argue for an assessment of this diluted effect since in order to know the effect of oral contraception on HIV risk based on how women actually take oral contraceptives (which may vary substantially by population). While others argue that an assessment of the diluted effect of oral contraceptives due to poor adherence is uninformative, since women are not being exposed (or not uniformly exposed) to exogenous hormones. In terms of measurement of exogenous hormones, this has not been conducted to assess reporting in analyses of HC and HIV risk. There were concerns over how informative this method could be since timing of the measurement would be very important. Additional research is needed on the possibilities of applying this measurement to HC-HIV studies. The presenters and meeting participants raised questions about how improved information about adherence would be utilized. For example, if a biologic assay for exogenous hormones is applied and a certain level of mis-reporting is observed, should this information be incorporated into the statistical model or should is only be used as adjunct information when interpreting the effect estimate. At present, there was no consensus as to how this information should be used.

- *Validation of sexual behavior reporting.* Improving measurement of sexual behavior is a challenging topic that could warrant its own meeting. In the context of this meeting, the presenters discussed several approaches that could be used to improve measurement of sexual behaviors. In terms of operational approaches, improving how questions are asked may improve sexual behavior reporting. Some studies have used computer-assisted self-interviews (CASI/ACASI) in an attempt to reduce misreporting to due social desirability. Mobile messaging (SMS) via cell phones could also be considered in future studies.

However, when using multiple measures, it is unclear which measure is the "right" one (i.e. best represents actual behavior). Analyzing a variety of different measures (both simple and complex) will improve our understanding of the robustness of these measures. In addition, it is important to assess whether the questions asked predict relevant biologic outcomes. Are rates of pregnancy, HIV, and STIs higher among women who report unprotected sex and highest among women who report the most unprotected sex? Measures that predict these biologic outcomes are likely to be more reliable than those that do not.

Biologic approaches may also improve our understanding of the reliability of self-reported sexual behavior data. For example, stored genital samples can be tested for prostate specific antigen (PSA) or Y-chromosome in order to measure recent sexual exposures without a condom (or with a condom that slipped or ruptured). To determine if reporting accuracy differs by HC use, PSA or Y-chromosome testing can be compared between HC users and users of non-hormonal methods among women who report 100% condom use (a study to evaluate this has been proposed by the presenters). Again, if there are inconsistencies in reporting detected by a biologic assay, further discussions are needed as to how this information should be integrated in the analysis.

- *Strategies to understand HIV exposure.* Some studies that have assessed the relationship between HC and HIV risk have enrolled women from the general population. A substantial fraction of those enrolled will likely have no HIV exposure during the course of follow-up. There is a potential for bias if HIV exposure is linked to contraceptive choice and/or condom use (i.e. if women who are more likely to use HC are also more likely to have partners that are HIV-infected). Women who know that their partner is HIV-infected may have different HIV risk perception, condom use, and HC use compared to women who know their partner in uninfected or don't know their partner's HIV status. Characterization of the level of HIV exposure in a population may be important and could be possible (outside of the context of serodiscordant couples). It could be argued that studies of general population women are a mixture of a) serodiscordant couples (recognized and unrecognized) and b) women not at risk. Testing of genital samples from women for HIV using DNA/RNA PCR to determine HIV exposure could help quantify HIV exposure. This has not been done previously and meeting participants raised concerns over the timing of testing and the need for validation of such methods. Further work in this area would be useful for improving our understating of HIV exposure in different populations.

- *Exploring effects by site.* Some studies have suggested differences in the effect of HC and HIV risk across site. It may be helpful to conduct sensitivity analyses by site, country or region to explore any potential differences. However, when planning and interpreting such analyses, it is important to consider the number of events in each sub-group and the power to detect associations within each sub-group.

Additional analyses and lab testing take time and resources, but might help to better understand the robustness of the primary results.

Summary of recommendations from meeting participants

After one and half days of thoughtful presentations and discussion about this important public health issue, it is clear that there is still much that is required in order to improve our understanding of the relationship between HC and HIV acquisition. A summary of key recommendations regarding the generation of new information, including future HC-HIV analyses using observational data and ancillary studies to inform primary analyses, is presented in Box 3. Additional thoughts were presented on several other topics. There was continued discussion regarding the tension between method sophistication (MSM) and whether the data collected will work for these methods, as missing and mis-reported data for key confounders and mediators may impede the implementation of certain methods. For those planning to use MSM, there was a call for the development of practical guidance on the implementation of MSM on other methods to assess the HC-HIV issue based on the challenges faced by others who have previously assessed this issue using these methods. In addition, the majority of recent and known planned HC-HIV analyses utilize data from HIV prevention trials. It was suggested that data sources from other fields be assessed as potential sources for future analyses.

Box 3. Key recommendations to generate new information that will improve our understanding of the effect of HC on HIV risk using observational data

Design and analytic approaches

- For secondary analyses using data from prospective cohorts or randomized trials, improve data collection in the parent study to allow for improved precision in ascertainment of HC exposure (with the ability to differentiate between similar methods, such as DMPA and NET-EN), potential confounders and mediators.
- When planning analyses, pay careful attention to the design. Recommend designing the analysis and comparison groups as if it were an RCT.
- Clearly define the comparison group and understand the implications of the choice of comparison group.
- Clearly define and acknowledge what is being assessed in the analysis (total effect versus direct effect not mediated by X) and the implications of information generated by the analysis.
- Utilize sensitivity analyses to assess the robustness of the findings, potential areas of bias, and the effect of different parameterization of potential confounders (i.e. sexual behaviors, condom use [and consistency of use]– are these factors predictive of biologic outcomes in the data?)

Ancillary study approaches

- There is a need for additional work to identify best practices to address pregnancy.
- There is a need to improve our understanding of the impact of missing data in these analyses and identify best practices to address missing data.
- There is a need to improve our understanding of the impact misclassification of key confounders and mediators in these analyses. Encourage the conduct of rigorous simulation studies using biological theories to assess how the different models we use can fix some of the problems in this data. How close we can get to the "truth" using simulated data to reflect what is going on in our real data? For example, how does misreporting of condom use affect the estimate?
- Conduct additional biologic studies to improve our understanding of if/how the probability of HIV exposure differs by choice of contraceptive method and how different levels of population HIV exposure affect estimates in these analyses.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the United States Agency for International Development.

Acknowledgements
We are grateful to USAID and the FHI360 PTA project for funding this meeting, and to the Bill and Melinda Gates Foundation for hosting. We thank all meeting participants, including: Jared Baeten (University of Washington), Jennifer Balkus (Fred Hutchinson Cancer Research Center/University of Washington), Matthew Barnhart (USAID/OHA), Diana Blithe (NIH/NICHD), Elizabeth Brown (Fred Hutchinson Cancer Research Center/University of Washington), Delivette Castor (USAID/OHA), Pai-Lien Chen (FHI360), Michael Chirenje (University of Zimbabwe), Steve Cole (University of North Carolina, Chapel Hill), Angela Crook (Medical Research Council, Clinical Trials Unit, UK), Kate Curtis (CDC/ONDIEH/ NCCDPHP), Deborah Donnell (Fred Hutchinson Cancer Research Center/University of Washington), Geoff Garnett (Bill and Melinda Gates Foundation), Emily Harris (USAID/OHA), Renee Heffron (University of Washington), Benny Kottiri (USAID/OHA), Jeanne Marrazzo (University of Washington), Sandi McCoy (University of California, Berkeley), Erica Moodie (McGill University), Charles Morrison (FHI360), Lisa Noguchi (Johns Hopkins Bloomberg School of Public Health), Sharon Phillips (World Health Organization), Chelsea Polis (USAID/PRH), Lauren Ralph (University of California, Berkeley), Denise Russo (NIH/NICHD/PAMA), Jim Shelton (USAID/PRH), Tabitha Sripipatana (USAID/PRH), David Stanton (USAID/OHA), James Trussell (Princeton University), and Daniel Westreich (Duke University). Finally, we are grateful to the study participants and investigators who have contributed to understanding of this complex subject.

References

1. Heffron R, Donnell D, Rees H, Celum C, Mugo N, Were E*, et al.* Use of hormonal contraceptives and risk of HIV-1 transmission: a prospective cohort study. *Lancet Infect Dis* 2012,**12**:19-26.
2. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S*, et al.* Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res* 2004,**4**:38.
3. Kavanagh BP. The GRADE system for rating clinical guidelines. *PLoS Med* 2009,**6**:e1000094.
4. Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med* 1965,**58**:295-300.
5. Rodriguez MI, Reeves MF, Caughey AB. Evaluating the competing risks of HIV acquisition and maternal mortality in Africa: a decision analysis. *BJOG* 2012,**119**:1067-1073.
6. Jain AK. Hormonal contraception and HIV acquisition risk: implications for individual users and public policies. *Contraception* 2012,**86**:645-652.
7. Butler AR, Smith JA, Polis CB, Gregson S, Stanton D, Hallett TB. Modelling the global competing risks of a potential interaction between injectable hormonal contraception and HIV risk. *AIDS* 2013,**27**:105-113.
8. Moodie EE, Stephens DA. Using Directed Acyclic Graphs to detect limitations of traditional regression in longitudinal studies. *Int J Public Health* 2010,**55**:701-703.
9. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology* 2009,**20**:18-26.
10. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000,**11**:550-560.
11. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000,**11**:561-570.
12. Moodie EE, Stephens DA. Marginal Structural Models: unbiased estimation for longitudinal studies. *Int J Public Health* 2011,**56**:117-119.
13. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008,**168**:656-664.
14. Hernan MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol* 2006,**98**:237-242.
15. Morrison CS, Richardson BA, Mmiro F, Chipato T, Celentano DD, Luoto J*, et al.* Hormonal contraception and the risk of HIV acquisition. *AIDS* 2007,**21**:85-95.
16. Morrison CS, Chen PL, Kwok C, Richardson BA, Chipato T, Mugerwa R*, et al.* Hormonal contraception and HIV acquisition: reanalysis using marginal structural modeling. *AIDS* 2010,**24**:1778-1781.
17. Chen PL, Cole SR, Morrison CS. Hormonal contraception and HIV risk: evaluating marginal-structural-model assumptions. *Epidemiology* 2011,**22**:877-878.
18. McCoy S, Zheng W, Montgomery E, Blanchard K, van der Straten A, de Bryn G*, et al.* Oral and injectable contraception use and risk of HIV acquisition amogn women in the methods for improving reproductive health in Africa (MIRA) study. *AIDS* 2013,**in press**.