Apples and oranges? Interpreting success in HIV prevention trials

Lori L. Heise1, Charlotte Watts1, Anna Foss1, James Trussell2, Peter Vickerman1, Richard Hayes3, and Sheena McCormack4

1Department of Global Health and Development, Social and Mathematical Epidemiology Group, London School of Hygiene and Tropical Medicine, London WC1H 9SH, UK

2Office of Population Research, Princeton University, Princeton NJ 08544, USA, The Hull York Medical School, University of Hull, Hull HU6 7RX, UK

3Department of Infectious Disease Epidemiology, MRC Tropical Epidemiology Group, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

4MRC Clinical Trials Unit, London NW1 2DA

Abstract

Background—In the last decade, several large-scale, clinical trials evaluating the efficacy of novel HIV prevention products have been completed, and 8 are currently underway or about to be reported. Little attention has been given in the literature to the level of protection sufficient to warrant introduction, and there is concern that using the term “efficacy” to describe the effect of user-controlled methods such as microbicides may mislead policymakers.

Design—We review how the fields of family planning, vaccine science, and mathematical modelling understand and use the terms efficacy and effectiveness and explore with simple mathematical models, how trial results of user-controlled products relate to common understandings of these terms.

Results—Each field brings different assumptions, a different evidence base, and different expectations to interpretations of efficacy and effectiveness—a reality that could cloud informed assessment of emerging data.

Conclusion—When making judgments on the utility of new health technologies, it is important to use standards that yield appropriate comparisons for the innovation and that take into account the local epidemic and available alternatives.

Keywords

efficacy; effectiveness; HIV prevention; vaccine; microbicide; contraception; family planning

© 2013 Elsevier Inc. All rights reserved.

Corresponding author: Lori L. Heise, Department of Global Health and Development, Social and Mathematical Epidemiology Group, London School of Hygiene and Tropical Medicine, London, W1H 9SH, lori.heise@lshtm.ac.uk.

Publisher’s Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
1. Introduction

Over the last decade, several large-scale, clinical trials evaluating the efficacy of various novel HIV prevention products, including experimental vaccines, use of oral HIV drugs for pre-exposure prophylaxis (PrEP), and a variety of new microbicide candidates for preventing HIV transmission to women during vaginal sex have been reported, and 8 more are underway or about to be reported [1]. The trial designs include a wide range of effect size for the sample size calculation (33%–60% reduction), but it is not clear what level of protection will be deemed sufficient to warrant introduction. This will depend in part on the how policy makers interpret the numbers emerging from clinical trials and how these compare to their expectations regarding efficacy and effectiveness. Given that condoms are routinely cited as 95% efficacious, for example, some HIV stakeholders perceive anything less as unacceptable.

In assessing “how good is good enough”, it is important to differentiate between user-controlled and provider-controlled methods, particularly in relation to the impact on estimates of biological efficacy that can be derived using trial data. There is also confusion in translating the results from trials into individual risks which are often easier for people to understand when deciding whether or not a product will work for them.

As a first step toward understanding these issues, we review the definitions and use of the terms efficacy and effectiveness across different fields, including vaccine science, HIV prevention, family planning, and mathematical modelling. Then we use mathematical modelling to illustrate how these various measures relate in the particular case of HIV transmission, and attempt to determine the appropriate standards for use for user-controlled methods.

2. Efficacy versus effectiveness

With respect to health outcomes, efficacy is the improvement achieved with use of an intervention by participants under ideal conditions. It frequently refers to research settings or situations of perfect use. By contrast, effectiveness is reserved for the effect that can be achieved in practice, taking into account limited coverage, constrained resources, and inconsistent or imperfect use.

Both efficacy and effectiveness are measures of relative rather than absolute risk; they compare the incidence of a health outcome among individuals receiving an intervention versus those who do not, and are expressed as ratios. The absolute risk by contrast is expressed as a simple proportion.

2.1. The meaning of efficacy and effectiveness in trial results

For interventions like vaccines, where adherence by trial participants can be measured objectively, it is often possible to get an accurate estimate of biological efficacy from the results of a Phase 3 trial, provided loss to follow-up is minimal (see Appendix 1 for how vaccine efficacy is calculated). This is generally based on a per-protocol analysis, restricted to those who complete the full immunization schedule (perfect use).

When evaluating products that are user-dependent, however, the trial results cannot be adjusted to derive efficacy unless there are reliable biological measures of product adherence. In trials of products such as microbicidal gels or PrEP, not all participants will use the product correctly or consistently, and measurement of use relies on self-reports. A Phase 3 trial therefore provides a combined measure of the product’s biological efficacy and the pattern of use in the trial (adherence). Indeed, because of incorrect and inconsistent use,
a trial that demonstrates a 40% reduction in incidence of HIV between its arms necessarily implies a product of higher efficacy.

2.2. Efficacy in the world of mathematical modelling

Modellers use the term “efficacy” differently to denote risk reduction for a single act of sexual intercourse with an infected partner rather than the protection achieved over time (as in trials). They use this per-sex-act efficacy in combination with many other factors—including the consistency with which it is used in different partnerships; the probability that the woman’s sexual partner is HIV-infected; whether he is in the high viremia phase; and whether either partner has other sexually transmitted infections—to project how use of a method may affect patterns of HIV transmission.

For outcomes that are relatively rare, like HIV, per-sex-act efficacy is a reasonable approximation of the efficacy of the method as understood by trialists or program planners. The same is not true, however, for more infectious pathogens, such as bacterial STIs, which have a higher transmission probability [2]. Thus, it is not always appropriate to equate the parameters used in modelling articles with the estimates of efficacy derived from clinical trials, although for HIV the values are comparable.

3. Current evidence of the efficacy of different HIV and contraceptive technologies

When the trial results for any new HIV prevention method become available, donors and policymakers will have to decide whether it merits further investment and ultimately introduction.¹ Their evaluation of the potential utility of any new method will depend in part on their assessment of the adequacy of existing options, their reading of available evidence, and the assumptions they may bring from past experience.

Current evidence on the efficacy and effectiveness of different prevention technologies varies widely. Likewise, the fields of HIV prevention, family planning/reproductive health, and vaccine science hold different working assumptions about what level of efficacy warrants introduction and use.

3.1. Efficacy of HIV prevention methods

In the world of HIV, the strongest evidence for an effective prevention strategy exists for male circumcision, which has been shown in 3 randomized controlled trials to reduce the risk of HIV acquisition among heterosexual men by roughly 50% to 60% [3–5]. The evidence for other HIV interventions is considerably less rigorous. For example, there have been no randomized controlled trials to assess the efficacy of the male and female condom against HIV and other STIs, although there have been several reviews of the available data [6,7]. Available estimates come from observational cohort studies of HIV discordant couples where a sero-negative person with a known exposure can be followed over time. When some individuals use condoms 100% of the time and some never use condoms, efficacy is calculated by taking one minus the ratio of the HIV incidence among “always users” of condoms versus “never users,” based on self-reports.

A Cochrane meta-analysis of these studies suggest a reduction in risk with consistent condom use of 80%, with the confidence limits for ‘consistent users’ ranging between 34%

---

¹Because many new HIV prevention technologies are being developed with public sector financing, the decision to take a product forward will lie with donors rather than a private company.
and 94% [8]. When inconsistent users are also considered, the condoms reduced the annual HIV incidence by 69% [9].

Reviews of condom effectiveness have also sought to estimate the per-sex-act efficacy of the condom. In particular, Pinkerton and Abramson [10] used a Bernoulli model to “back calculate” the likely per-sex act efficacy of male condoms associated with observed reductions in incidence among discordant couples. They concluded that condoms reduced the probability of HIV transmission per-sex-act by as much as 95%. It is this, per-sex-act, modelled figure that is often cited for both condom efficacy and effectiveness (see Table 1).

3.2 Efficacy of contraceptive methods

Since many women receive most of their sexual and reproductive health care from family planning and pre-natal clinics, these sites are likely to be among the venues used to distribute any new HIV method aimed at women in the general population. Thus, family planning and reproductive health providers are likely to be among the various “gatekeepers” whose opinion will shape how women perceive any new HIV prevention method.

Family planning providers have been socialized in a world where contraceptive efficacies of 96 to 99% are not uncommon and anything less than 90% is largely dismissed as inadequate. After a decades-long effort to develop highly effective, long-acting contraceptive technologies (such as intrauterine contraceptives, injectables, and implants), the benchmarks for contraceptive efficacy are now quite high (see Appendix 1 for equations of how contraceptive efficacy is calculated).

Estimates of contraceptive efficacy and effectiveness, however, vary in rigor. Evidence of efficacy for most methods comes from clinical investigations, but there have been no randomized, placebo-controlled trials of modern contraceptives. Sometimes clinical studies have randomized women to use different methods, but in many cases estimates derive from observational cohort studies, which track unwanted pregnancies among women using different birth control methods over the course of (usually) one year [11]. Like other observational studies, such research is subject to bias.

4. Explaining partial protection

Regrettably, data from trials do not translate easily into concepts accessible to individuals wishing to assess their own risk or to policymakers and donors seeking to make investment decisions among options. Research shows that individuals have a difficult time understanding notions of probability, especially when information is presented in relative, rather than absolute, terms, as is the case with trials [12]. Relative measures are especially problematic because most people do not have an internalized notion of the underlying probabilities: for example, 42% of unmarried men and 40% of unmarried women aged 18–29 years in the United States believe that the chance of getting pregnant within a year while using the birth control pill is 50% or greater [13].

To compensate, the family planning field tends to rely instead on a measure known as the probability of contraceptive failure. Unlike efficacy and effectiveness, which represent a relative reduction in risk of unwanted pregnancy (a ratio), this provides women and providers with an estimate of the absolute probability that a woman will get pregnant if she uses a method over a period of a year (a proportion).

When discussing contraceptive options, providers commonly present two different pieces of information: probabilities of failure during a year of ‘perfect use’ and during a year of ‘typical use’ (including inconsistent or incorrect use) [14]. This metric has been shown to be

*Contraception. Author manuscript; available in PMC 2013 September 17.*
far easier for women to interpret, because it helps them assess their personal likelihood of becoming pregnant within a time frame that is easy to grasp.

In general, there are large differences between probabilities of contraceptive failure during perfect use and during typical use for methods that require adherence, and smaller differences for less user-dependent products. For example, the annual risk of an unwanted pregnancy among women who use male condoms consistently and correctly is only 2% (perfect use) but it is 15% among typical users (see Table 1) [14].

Unfortunately, this approach is more difficult to take in the world of STIs and HIV, where the risk of HIV acquisition is largely defined by the infection status of one’s sexual partner. Not only does the risk of infection per act of intercourse vary dramatically by pathogen but the risk of being "exposed" also varies dramatically by partner. Thus, it is much harder to construct a metric for women to understand their absolute risk of HIV infection over a year because it is defined as much by who they have sex with as it is by what prevention method they may use.

Likewise, with user-dependent methods like microbicides or condoms, focusing on the method’s efficacy alone misses half the story. The protection that a prevention method confers is a function of both the inherent efficacy of the method and how consistently it is used. Indeed, given the particular transmission dynamics of HIV, consistency directly compensates for efficacy [15]. In other words, using a low efficacy method consistently for HIV protection can confer as much protection as using a high efficacy method inconsistently.

To illustrate this, Table 2 shows model projections of how a woman’s absolute risk of acquiring HIV varies, depending upon the likelihood that her partner is HIV-infected, and the extent to which she can use different methods of protection. Using simple model projections of the probability of HIV acquisition over a year, this model projects that for the scenarios presented, and using HIV transmission probabilities appropriate for developing country settings [16], women in discordant partnerships have a one in five risk of acquiring HIV over a year if they do not use condoms, which drops to a risk of 0.2% for women in lower risk partnerships (2 in 1000 risk).

Assuming that microbicide use reduces the per-sex-act risk of HIV acquisition by 60%, and condoms reduce the per-sex-act probability of HIV acquisition by 95%, we can see that different combinations of product use can achieve similar levels of protection. For example, if there is a 50/50 chance that a woman’s next sexual partner is likely to be infected, based on the underlying HIV prevalence in her sexual network, she can achieve the same level of protection through 50% condom use or 75% microbicide use (6.4% vs. 6.2%).

5. What do trial results tell us about product efficacy?

So, what do the relative risk numbers emerging from clinical trials tell us about product efficacy? That depends on the nature of the product and on how the results are reported. Because vaccines are administered by providers, Phase 3 clinical vaccine trials can yield true estimates of efficacy. Among participants who receive all doses, breakthrough infections more clearly reflect product failures rather than failure to use the product.

With interventions that require user adherence, Phase 3 trials do not yield true estimates of efficacy; rather they provide an unbiased measure of effect in the trial setting, but one that
represents an average across all users, including those who did not use the product or did so inconsistently. Nonetheless, it is possible to estimate product efficacy from the results of an effectiveness trial of a user-dependent product. Algebraic analysis suggests that a product’s efficacy can be approximated by dividing the measured trial effectiveness by the consistency of gel use in the non-condom protected sex acts (see Appendix 1 for supporting math).

Table 3 provides examples of the imputed efficacy that would be implied by relatively modest trial results of 30 to 40% reduction in acquisition of HIV among comparison groups. Depending on the level of product adherence achieved in the trial, the imputed product efficacy could range from 33 to 57% (or higher if reported adherence is lower than 70% among women not using condoms). It is this efficacy value that is most comparable with the 95% efficacy value cited for condoms.

While adherence is admittedly subject to over-reporting, any overestimation of use will lead to a conservative, rather than inflated, estimate of product efficacy. [The higher the reported adherence, the lower the estimate of product efficacy]. By way of comparison, the reported gel adherence in non-condom protected acts in the recent HPTN 035 trial of the Pro2000 microbicide gel was 68.2% (Barbara Richardson, Fred Hutchinson Cancer Research Center, personal communication). Among women in the MDP301 study who attended all 13 study visits, 56% reported using gel at every last sex act.

6. Interpreting trials: How good is good enough?

Given the different understandings of the terms efficacy and effectiveness, different stakeholders are likely to answer this question differently. Those coming from the world of preventive vaccines routinely accept vaccines that are 70% efficacious, including vaccines against pertussis, hepatitis B for newborns, and polio [17–19]. In family planning, the “gold standard” is considerably higher because there are many forms of contraception that achieve 95% to 99% efficacy in clinical studies (although in no study were women randomized to receive a placebo). But does a similarly high bar make sense for HIV where there are still groups of users who have no method of HIV protection they can reliably implement?

One such group are women in regular partnerships. Studies from around the world have shown that couples, especially women, find it extremely difficult to negotiate consistent condom use within ongoing relationships. Yet, it is precisely within regular partnerships that many women are becoming infected [20,21]. While the female condom is theoretically available, in reality it remains scarce, expensive, and ill-suited for relationships where women want to have children. By contrast, both women and men have responded enthusiastically to the use of microbicide gels within regular partnerships, with a significant proportion reporting increased sexual pleasure [22]. Moreover, mathematical modelling has shown that even a moderately effective microbicide could have a significant impact on HIV transmission in some settings [23]. Similar impacts have been demonstrated for partially effective HIV vaccines [24,25].

Nonetheless, many users and policymakers have internalized the belief that condoms reduce the risk of HIV acquisition by 95%, leaving the impression that this is the level of effectiveness against which any future HIV prevention methods should be compared. But this 95% figure is the estimated protection condoms provide in a single act of intercourse (not over time), derived from mathematical modelling. It is more appropriate to compare the results of trials of user-controlled methods such as microbicides, to the 69% reduction observed in the meta-analysis of consistent and inconsistent condom users over time.

Assessing what is “good enough” also depends on the underlying risk of disease. In settings where one in three sexually active adults are infected with HIV, as is the case in some parts...
of sub-Saharan Africa, a method that halves risk will be received differently than in settings where fewer than 1 in 1000 people are infected.

Likewise, perceptions will be shaped by how data are presented to women and to policymakers. Spermicides are 90% efficacious at preventing pregnancy, but among 100 women using spermicides for birth control for 12 months, 29% will have an unwanted pregnancy. These data describe the same product, but the first is a relative risk estimate comparing consistent users to non-users while the second communicates the absolute risk of becoming pregnant in a year among typical users [14].

Inconsistency between different types of trial analyses [intent-to-treat (ITT) vs. per protocol] can be a further cause of confusion, as demonstrated by the recent release of the results from Thai RV144 HIV vaccine study. Investigators from this trial first released to the press results from the trial’s primary endpoint—a modified ITT analysis that included everyone who was randomized (even if they did not receive all of the trial’s required injections), but excluded 7 individuals found to be HIV-positive at baseline. They later released a per-protocol analysis that included only participants who were HIV-negative at baseline and who received all six injections. The modified ITT analysis found statistically significant protection (31.2% reduction; p=0.035), whereas the per-protocol analysis did not (26.2% reduction; p=0.16). The results were not numerically inconsistent with each other, but the ITT result reached significance because of the larger number of endpoints, and this created a sense of inconsistency [26].

At times ITT and per-protocol results can diverge substantially. The Merck HPV vaccine, for example, is widely promoted as 99% efficacious (as revealed in the per-protocol analysis), but this efficacy estimate refers only to prevention of dysplasia and/or cancers caused specifically by HPV strains 16 and 18 among women who received all three doses of the vaccine. In an ITT analysis (which includes women who received less than 3 doses), the vaccine was 44% efficacious against HPV 16- and 18-related negative outcomes and only 18% efficacious against all HPV-related dysplasia or cancer [27]. Our point is not to question the public health importance of the HPV vaccine, but rather to highlight how confusing different analyses can be to the untutored eye.

Thus, when making judgments on the utility of new health technologies, it is important to use appropriate benchmarks and to compare innovations against available alternatives. Policymakers must be careful not to mix apples and oranges: they must separate trial results from extrapolated modelling estimates; consider per-protocol analysis and ITT analysis within the context of user-controlled and provider-controlled methods; and avoid setting inappropriate standards for HIV risk reduction based on metrics created for family planning. People’s future access to an expanded array of prevention options will depend in part on our collective ability to draw appropriate and informed comparisons.

References


Contraception. Author manuscript; available in PMC 2013 September 17.

Table 1

Estimates of condom efficacy and effectiveness

<table>
<thead>
<tr>
<th></th>
<th>HIV prevention</th>
<th>Contraception</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Always users vs. never users (efficacy)</td>
<td>Consistent &amp; inconsistent users vs. never users (effectiveness)</td>
</tr>
<tr>
<td>Efficacy or effectiveness</td>
<td>80%</td>
<td>69%</td>
</tr>
<tr>
<td>Annual probability of failure</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Sources: [8,14]
Model estimates of the annual probability of a woman acquiring HIV in one year, assuming that a microbicide reduces the per sex act risk of HIV acquisition by 60%, for different patterns of product use and assumptions about the likelihood that her sexual partner is HIV-infected.

<table>
<thead>
<tr>
<th>Probability partner HIV-infected</th>
<th>100</th>
<th>75</th>
<th>50</th>
<th>25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>No product use</td>
<td>21.0</td>
<td>15.7</td>
<td>10.5</td>
<td>5.2</td>
<td>0.2</td>
</tr>
<tr>
<td>20% condom use, no microbicide use</td>
<td>17.7</td>
<td>13.3</td>
<td>8.9</td>
<td>4.4</td>
<td>0.2</td>
</tr>
<tr>
<td>0% condom use, 75% microbicide use</td>
<td>12.9</td>
<td>9.6</td>
<td>6.4</td>
<td>3.2</td>
<td>0.1</td>
</tr>
<tr>
<td>50% condom use, no microbicide use</td>
<td>12.4</td>
<td>9.3</td>
<td>6.2</td>
<td>3.1</td>
<td>0.1</td>
</tr>
<tr>
<td>20% condom use, 75% microbicide use</td>
<td>10.8</td>
<td>8.1</td>
<td>5.4</td>
<td>2.7</td>
<td>0.1</td>
</tr>
<tr>
<td>50% condom, 75% microbicide use</td>
<td>7.5</td>
<td>5.7</td>
<td>3.8</td>
<td>1.9</td>
<td>0.1</td>
</tr>
</tbody>
</table>

 Assumes 1 sexual partner, 100 sex acts per year, 5% of HIV-infected partners have increased risk of transmission due to high viremia, condoms reduce per sex act transmission risk by 95%, male-to-female HIV transmission probability 0.193%.

More information on the modeling behind this table is provided in Appendix 1.
Table 3
Imputed product HIV efficacy, for different assumptions about the measured trial effectiveness, and levels of gel adherence in non-condom protected sex acts

<table>
<thead>
<tr>
<th>Measured effectiveness in trial</th>
<th>Gel adherence in non-condom protected sex acts</th>
<th>Estimated product efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>30%</td>
<td>70%</td>
<td>43%</td>
</tr>
<tr>
<td>30%</td>
<td>80%</td>
<td>38%</td>
</tr>
<tr>
<td>30%</td>
<td>90%</td>
<td>33%</td>
</tr>
<tr>
<td>40%</td>
<td>70%</td>
<td>57%</td>
</tr>
<tr>
<td>40%</td>
<td>80%</td>
<td>50%</td>
</tr>
<tr>
<td>40%</td>
<td>90%</td>
<td>44%</td>
</tr>
</tbody>
</table>

Supporting math available in Appendix 1, on-line.