



# HHS Public Access

Author manuscript

*Contraception*. Author manuscript; available in PMC 2016 January 04.

Published in final edited form as:

*Contraception*. 2011 August ; 84(2): 184–193. doi:10.1016/j.contraception.2010.11.012.

## Quantitative perceptual differences among over-the-counter vaginal products using a standardized methodology: implications for microbicide development<sup>☆</sup>

Ellen D. Mahan<sup>a,b</sup>, Kathleen M. Morrow<sup>c</sup>, and John E. Hayes<sup>a,b,\*</sup>

<sup>a</sup>Sensory Evaluation Center, College of Agricultural Sciences, The Pennsylvania State University, University Park, PA 16802, USA

<sup>b</sup>Department of Food Science, College of Agricultural Sciences, The Pennsylvania State University, University Park, PA 16802, USA

<sup>c</sup>Department of Psychiatry and Human Behavior, Warren Alpert Medical School, Brown University, Providence, RI 02912, USA

### Abstract

**Background**—Increasing prevalence of HIV infection among women worldwide has motivated the development of female-initiated prevention methods, including gel-based microbicides. User acceptability is vital for microbicide success; however, varying cultural vaginal practices indicate multiple formulations must be developed to appeal to different populations. Perceptual attributes of microbicides have been identified as primary drivers of acceptability; however, previous studies do not allow for direct comparison of these qualities between multiple formulations.

**Study Design**—Six vaginal products were analyzed *ex vivo* using descriptive analysis. Perceptual attributes of samples were identified by trained participants ( $n=10$ ) and rated quantitatively using scales based on a panel-developed lexicon. Data were analyzed using two-way ANOVAs for each attribute; product differences were assessed via Tukey's honestly significant difference test.

**Results**—Significant differences were found between products for multiple attributes. Patterns were also seen for attributes across intended product usage (i.e., contraceptive, moisturizer or lubricant). For example, Options<sup>®</sup> Gynol II<sup>®</sup> (Caldwell Consumer Health, LLC) was significantly stickier and grainier than other products.

**Conclusions**—Descriptive analysis, a quantitative approach that is based on consensus lexicon usage among participants, successfully quantified perceptual differences among vaginal products. Since perceptual attributes of products can be directly compared quantitatively, this study represents a novel approach that could be used to inform rational design of microbicides.

<sup>☆</sup>This manuscript was completed in partial fulfillment of the requirements for a Master of Science degree at the Pennsylvania State University by EDM. The project was supported by faculty startup funds to JEH.

<sup>\*</sup>Corresponding author. Pennsylvania State University, University Park, PA 16802, USA. Tel.: +1 814 863 7129. jeh40@psu.edu (J.E. Hayes).

## Keywords

Vaginal creams; foams and jellies; Acceptance; Topical microbicides

---

## 1. Introduction

### 1.1. Need for female-focused HIV prevention

While the increasing prevalence of HIV and AIDS has long been recognized as a worldwide epidemic, recently more awareness and priority have been given to HIV prevalence among heterosexual women [1,2], including the need for women-initiated and -controlled prevention options like microbicides [3]. As of June 2010, there were at least seven different microbicide formulations in Phase 1–3 clinical trials in planning or underway [4]. Although prior attempts to reduce HIV transmission via microbicides had been unsuccessful, this approach was recently validated by new evidence that microbicides can successfully reduce HIV transmission rates in women [5]. The success of microbicides from a formulation standpoint relies on several factors: safety, efficacy, chemical/physical stability, access and affordability, distribution and retention in the vagina, and user acceptability [6,7].

### 1.2. Acceptability as a potential driver of effectiveness

Acceptability of a microbicide is extremely important for microbicide success: a product could be a very potent inhibitor of HIV infections but if a woman will not use it, the product is useless. Acceptability is multifactorial and depends on properties such as packaging, side effects, safety and perceptual properties [7,8]. These perceptual properties can include the appearance, smell and taste of the microbicide or texture-based properties such as impact on sexual pleasure (how the product feels during intercourse) or leakage (the propensity of the product to seep out of the body). Common terms used in acceptability trials to describe negative aspects of candidate microbicides include ‘messy’, ‘drippy’ and ‘sticky’ [9,10]. However, due to the nature of acceptability trials, only one microbicide formulation is typically evaluated at a time and, as such, attributes among formulations have not been directly compared. Other industries have approached direct attribute comparisons using a technique called descriptive analysis (e.g., Ref. [11]).

### 1.3. Connecting formulation, sensory properties and consumer acceptability

Descriptive analysis has been utilized by the food and consumer products industries since the middle of the last century [12,13]. These methods assume that a panel of individuals can be trained to describe and reliably quantify product perceptions. Recruitment criteria are based on ability to detect small differences between products, describe differences verbally and use scales accurately. This is followed by training, where a panel leader facilitates lexicon development, which participants use to describe attributes of the products. Subsequently, scales, anchors and references are also identified. Quantitative data are collected blindly and independently and then analyzed statistically. The goal is to get consistent and reliable quantitative data that describe perceptual attributes of a product [12,14]. Traditionally applied to food products, these methods are not limited solely to foods [11].

## 1.4. Goals and hypothesis

Schwartz [15], a scientist at General Foods, first adapted descriptive analysis to skin care products in 1974. Subsequently, the American Society for Testing and Materials (ASTM E1490-03; available from <http://www.astm.org/Standards/E1490.htm>) developed a standardized method to evaluate skin care products. Here, we apply similar methods to six over-the-counter (OTC) vaginal products. Although others have evaluated commercial products for anti-HIV activity [16] or used commercial products as surrogates for microbicides in acceptability studies [17], we are unaware of published quantitative data on perceptual differences between vaginal products. Here, we ask whether a trained panel, using standard methods adapted from sensory science, can discriminate quantitative differences among OTC vaginal products when evaluated on the skin *ex vivo*.

## 2. Methodology

### 2.1. Samples

The six OTC products evaluated included three personal lubricants, KY<sup>®</sup> Jelly (Johnson & Johnson), Pre-Seed<sup>®</sup> (INGfertility<sup>™</sup>) and Astroglide<sup>®</sup> (Biofilm, Inc); two personal moisturizers, Replens<sup>®</sup> (Lil' Drug Store Products, Inc.) and RepHresh<sup>®</sup> (Lil' Drug Store Products, Inc.); and one contraceptive gel, Options© Gynol II<sup>®</sup> (Caldwell Consumer Health, LLC), which contains a low level of the spermicide nonoxynol-9. Samples were stored in retail packaging at room temperature and all packages used were obtained from the same lot numbers.

### 2.2. Participants

Participants, recruited via email from an internal database maintained by our laboratory of interested individuals, were screened for age, sex, availability, tactile acuity, having no known skin allergies and not pregnant or nursing. Tactile acuity was screened using sandpaper swatches (described below). Women aged 18–40 years were invited to participate; three participants dropped out due to scheduling conflicts and 10 participants completed the entire study. As panelists are aligned to the individual attributes (see Discussion) and products undergo replicated assessment, this panel size is typical and is sufficient to reveal differences across products [18]. Of those who completed the study, four had previously participated in descriptive panels; however, none had participated in a panel involving topically applied products. Participants were aware of the broad product category but were blinded to the specific products used. A confidential follow-up questionnaire indicated that 50% had prior experience with these types of products. All procedures were approved by the Pennsylvania State University Institutional Review Board. Participants were compensated for their time.

**2.2.1. Screening for tactile acuity**—For tactile acuity screening, participants were required to complete four triangle tests, of varying difficulty, using sandpaper swatches of different grits; samples were covered with a paper flap to prevent visual assessment. In a triangle test, three samples are presented; of these, two are the same while one is different. Participants identify the odd sample. We screened 27 individuals; eight got all four tests correct and 10 got three of four correct. Of these 18, we invited 13 to participate, based on

availability and prior panel experience. At chance, the odds of getting three triangles tests correct is 3.7% ( $1/3 \times 1/3 \times 1/3$ ), indicating the selected panelists were able to discriminate samples well above chance performance and were not merely guessing.

### 2.3. Procedure overview

The panel was trained using descriptive analysis based on the ASTM E1490-03 guidelines. Training began with lexicon development and identification of reference standards followed by panel calibration and quantitative evaluation using Compusense<sup>®</sup> five software (Compusense, Inc., Guelph, ON, Canada). All samples were dispensed using 3-mL clear plastic syringes (BD Luer-Lok<sup>™</sup>) labeled with three-digit blinding codes. Tactile evaluation consisted of three phases: petri dish, between the fingers and on the forearm (Fig. 1). Before evaluation, participants wiped fingers and forearms with a baby-wipe and three non-overlapping 4-cm circles were drawn on the right inner forearm; all circles were at least 1 in. above the wrist and 1 in. below the elbow to ensure the same type of skin was used for all assessments.

**2.3.1. Lexicon development and training**—Participants were trained as a panel in seven 1-h sessions over a 3-week period. Initially, participants were presented with one sample at a time and asked to generate terms describing various attributes of the samples. As terms were developed and more samples were evaluated, participants were encouraged by the panel leader to eliminate redundant terms and come to a consensus on terms that describe important attributes. This process is referred to as lexicon development; the final result was a list of attribute descriptors, definitions, anchors and references that were used during quantitative evaluation (see Table 1). References are products outside of the evaluation set used to demonstrate extremes of each descriptor to give panelists a contextual framework during sample evaluation. For example, for the descriptor ‘thickness’, the low-end reference (anchor: thin) was mineral oil and the high-end reference (anchor: thick) was lanolin. Training was also used to ensure that panelists manipulated samples in the same manner. Because a standard methodology has not been published previously for vaginal product evaluation, the panel leader guided the panel in reaching a consensus on how the products should be manipulated and evaluated in each phase (petri dish, fingers and forearm). The final ballot was generated using descriptors and manipulations that were decided upon during lexicon development.

While lexicon development is a consensus process, actual evaluations are performed independently. Panelists practiced evaluating samples in isolated evaluation booths during training to become calibrated to the scales and attributes. Additionally, we used the Feedback Calibration Method (FCM) option in Compusense<sup>®</sup> five, which has been shown to calibrate panels in an unbiased manner and reduces the time required to train a panel [19]. The goal of this process is to be able to use humans as calibrated sensors via a standardized methodology.

**2.3.2. Quantitative evaluation**—All evaluations took place in individual evaluation booths under white light using Compusense<sup>®</sup> five to collect data. Panelists were presented with a sample tray containing three syringes, three petri dishes, and baby-wipes. A

metronome set at 120 beats per minute (bpm) was played in the booths to standardize manipulation rates. All attributes were evaluated on continuous line scales labeled at each end with the appropriate anchor. Quantitative evaluation occurred over six sessions with each participant evaluating three products per session. Samples were presented in a randomized design; each participant evaluated all six products and each evaluation was replicated three times. Replicates were averaged prior to statistical analysis.

**2.3.3. Petri dish**—For petri dish evaluation, 0.5 mL of the sample was dispensed into the center of a clear glass petri dish marked with a 4-cm-diameter circle. Participants observed product in the dish and evaluated ‘whiteness’ and thickness after tilting the dish twice at a 45° angle. The dish was then set down and participants lightly tapped with the left index finger to evaluate ‘peaking’ and ‘ropiness’. The sample was then rubbed in a circular motion within the confines of the drawn circle at two rotations per second (rps) for 15 s and then for an additional 45 s, for a total manipulation time of 60 s. At 15 and 60 s, participants evaluated ‘smoothness’ and ‘air bubbles’. ‘Graininess’ was evaluated only at 15 s as panelists felt graininess did not change with manipulation during training. The sample was then spread over the entire surface of the dish to evaluate ‘uniform thickness’. The sample was allowed to sit for 1 min in the dish and tapped to evaluate ‘stickiness’.

**2.3.4. Fingers**—Participants were instructed to wipe fingers with a baby-wipe and paper towel and then dispense 0.1 mL between the index finger and thumb on the left hand. Participants first compressed the finger and thumb together and opened them slowly; then they compressed again, opening their fingers quickly. Following each compression (slow and fast), stickiness, peaking and ropiness were evaluated. Fingers were compressed once more and ‘rubberiness’ was evaluated. Participants then rubbed their fingers back and forth at the rate of two beats per second (bps) for 15 s and then an additional 45 s, for a total manipulation time of 60 s. Thickness, smoothness and slipperiness were evaluated at 15 and 60 s. After the 60 s of manipulation, participants opened fingers slowly and once again evaluated stickiness, peaking and ropiness.

**2.3.5. Forearm**—Fingers were again wiped with a baby-wipe and paper towel and participants dispensed 0.1 mL of the sample into the center of the circle previously drawn on the forearm. The sample was rubbed in a circle using the index finger at the rate of 2 rps for 15 s and followed with another 45 s. At time points 15 and 60 s, thickness, smoothness, slipperiness and air bubbles were evaluated. Following this manipulation, ‘amount left’ was evaluated.

## 2.4. Statistical analyses

Data were analyzed using two-way (Product×Participant) main effects ANOVAs for each attribute across products. Consistent with accepted practices [12], attributes were considered separate hypotheses, so no multiple comparisons adjustment was required across attributes. The participant factor in the ANOVA was significant in most cases, but is not reported here as it was not of theoretical interest. Although the purpose of FCM training is to calibrate panelists, some individuals may show some small idiosyncrasy in scale usage (e.g., they prefer the high or low end of the scale). Because the ultimate goal is to identify relative

differences among products, partitioning out panelist as a source of variance in the ANOVA enhances the ability to find product differences. Within an attribute, significant differences in product means were tested via Tukey's honestly significant difference test.

### 3. Results

In general, remarkable consistency was found among the three different assessment methods (e.g., petri dish, fingers and forearm), so the results are presented by attribute for simplicity and clarity. To facilitate comparisons across products and attributes, spider plots (Figs. 2 and 3) are typically used in sensory science, as this leverages the ability of the visual system to detect patterns across products. Fig. 4 summarizes the same data with conventional bar graphs and provides *F* values and significant differences between product means.

#### 3.1. Thickness (petri dish, fingers, forearm)

Regardless of the evaluation phase, the relative rank ordering with respect to thickness was Gynol>Replens≈RepHresh>KY>Astroglide≈Pre-Seed (where '>' indicates a significant difference and '≈' means approximately equal). *F* values and significant differences between product means are shown in Fig. 4.

#### 3.2. Smoothness (petri dish, fingers, forearm)

Smoothness was evaluated on a scale of smooth to clumpy; Gynol was the clumpiest sample, while all other samples were quite smooth. Across all evaluation phases, the relative rank ordering from least to most smooth was Gynol>KY>Replens≈RepHresh>Pre-Seed>Astroglide.

#### 3.3. Slipperiness (fingers and forearm)

Across all evaluation phases (fingers and forearm), the relative rank ordering from most to least slippery was Astroglide>Pre-Seed≈KY>Replens≈RepHresh>Gynol.

#### 3.4. Stickiness (fingers, petri dish)

Stickiness was evaluated before and after manipulation (i.e., rubbing the fingers for 60 s at 2 rpm) between the fingers and in the petri dish. Across all evaluation methods, the relative rank ordering from most to least sticky was Gynol>KY≈RepHresh>Replens>Astroglide≈Pre-Seed.

#### 3.5. Peaking (petri dish, fingers before and after rubbing)

As above, peaking was evaluated before and after manipulation (i.e., rubbing the fingers for 60 s at 2 rpm) and in a petri dish. After manipulation, all products showed markedly less peaking, although differences were still observed across products. The petri dish and fingers were not in complete agreement as RepHresh showed more peaking than Gynol ( $p<.05$ ) in the petri dish. In summary, Pre-Seed and Astroglide showed minimal peaking, but other results disagreed somewhat across methods.

### 3.6. Ropiness (fingers before and after rubbing, petri dish)

Astroglide was much ropier than all others. Also, the finger opening speed appeared to matter for ropiness, as fast opening enhanced the ability to discriminate between products ( $F$  value 529.8 vs. 346.9 for fast vs. slow opening, respectively). Nonetheless, across all evaluation phases, Astroglide was much ropier than KY, which was ropier than Replens, Gynol, RepHresh and Pre-Seed, which were all minimally ropy.

### 3.7. Air bubbles (forearm and petri dish)

Initially, on the forearm, the rank order for air bubbles was Gynol>Replens≈KY>RepHresh≈Astroglide≈Pre-Seed. However, after rubbing for an additional 45 s on the forearm, the amount of bubbles in Astroglide increased so that it moved from fifth to second in relative rank order, although it was still much lower than Gynol (5.2 vs. 1.6;  $p<.05$ ). In summary, bubble ratings on glass were higher than on skin, Gynol had many more bubbles across all conditions, Replens bubbles were high on glass but low on skin, Astroglide bubbles increased with additional manipulation on skin (but not glass) and the amount of bubbles in RepHresh and Pre-Seed were consistently low.

### 3.8. Other attributes

Regarding whiteness, Replens was much whiter than any other sample (9.0 vs. 5.7 for Gynol) and Gynol was much whiter than the remaining four products (5.7 vs. below 0.8). Other samples were essentially clear, although Pre-Seed was whiter than Astroglide (0.8 vs. 0.2).

Uniform thickness was evaluated visually after spreading the product across the petri dish. Gynol spread less evenly than KY, Replens and RepHresh, which did not differ from each other. Astroglide and Pre-Seed spread more evenly than the remaining four products.

Differences in graininess were observed when evaluated in the petri dish. Gynol was the only sample that was found to be grainy ( $p<.05$ ); all other samples received scores indicating that they were not grainy.

Fingers were compressed together to evaluate rubberiness, revealing differences across products. Gynol was the most rubbery sample followed by KY and then RepHresh and Replens, which did not differ. Pre-Seed and Astroglide were not rubbery.

When evaluating the amount of product left on the forearm after manipulation, Astroglide left more product on the arm than KY, the next highest sample (71% vs. 57%). The remaining five products (range 57–48%) did not differ significantly from each other. During evaluation, all products stayed wet; none became dried or flaky.

### 3.9. Summary of results

One major pattern that emerged was differences across intended product usage. Specifically, Gynol, the only contraceptive gel among our products, had a dramatically different profile across numerous attributes. Gynol was thicker, stickier, grainier, more rubbery, had more air bubbles, was the least slippery, least smooth and had the most uneven spread. The two

vaginal moisturizers, Replens and RepHresh, had very similar properties and only differed significantly from each other for the attributes stickiness, air bubbles and whiteness (RepHresh is stickier, while Replens has more air bubbles and is white). Notably, these products are manufactured by the same company and contain similar ingredients. Two lubricants, Pre-Seed and Astroglide, were very similar on several attributes. Both products spread evenly, were very slippery (Astroglide>Pre-Seed), thin, smooth and clear, and showed minimal peaking, air bubbles, stickiness, graininess and rubberiness. In contrast, KY, the other lubricant, was more similar to the vaginal moisturizers. Compared to other lubricants, KY was thicker, stickier, more rubbery and showed peaking and air bubbles; it did not spread as evenly as Pre-Seed or Astroglide either. However, KY was as slippery as Pre-Seed and was the only product besides Astroglide that was ropy.

#### 4. Discussion

In practice, vaginal microbicides have been well received by women of several different cultural backgrounds. Research involving women in the United States [20], Malawi [21], Zimbabwe [22,23] and China [24] has shown high acceptance of microbicide products among study populations. A 2005 review on microbicide acceptability indicated that, at the time, 77% (41 out of 53) of acceptability studies collected data on physical characteristics of products including odor, color, texture and viscosity [25]. Here, we sought to quantify, compare and, to a degree, define perceptual characteristics of vaginal products using humans as sensors in a standardized methodology.

The present work demonstrates that classical sensory science methods have utility in discriminating between vaginal products *ex vivo* as they identified large differences across multiple perceptual attributes. These methods also have the potential to inform the rational design of candidate microbicides. For example, novel formulations could be evaluated perceptually prior to safety testing. If integrated with acceptability data, this can streamline development by identifying formulations that should not proceed to the next phase (i.e., stage gating), as they are unlikely to be acceptable based on prior data. Slipperiness is discussed below as an example.

Focus group participants have used similar descriptors to those generated by our descriptive panel here. In an acceptability trial of Pro-2000 (a water-based investigational microbicide), participants described the gel as ‘smooth’, ‘thick’ and ‘not sticky’; these terms were viewed as positive attributes of the product [10]. However, ‘sticky’ and ‘slippery’ have been used by focus group participants to indicate both positive and negative aspects of vaginal gels [20,22,26]. For instance, slipperiness, associated with lubrication during sex, is desirable in cultures where ‘wet’ sex is expected. Women experiencing vaginal ‘dryness’ may prefer a more ‘slippery’ product [9] but enhanced lubrication is not always desirable, depending on cultural norms [27,28] or personal preference. Critically, this suggests that the field should work toward multiple formulations to meet varied user preferences as it is unlikely that a single microbicide will be optimal for all users globally.

A focus group conducted by Zubowicz et al. [17] used Replens as a microbicide surrogate to gauge reactions of adolescent women. Participants felt enhanced lubrication caused

‘messiness’ or made sex ‘too wet’. To prevent this, some reported using less product than recommended; this has critical implications for dosing of active ingredients in microbicides. Specifically, efficacious agents may still fail in the field if acceptability alters correct utilization.

One of the three ways descriptive analysis differs from previous focus group methods is the generation of quantitative data that can be analyzed with parametric statistics. Given that we are using the same participants to quantitatively evaluate multiple products, we can directly compare attributes across products. We found that Replens was significantly less slippery than KY, Pre-Seed and Astroglide. Since the Zubowicz participants found Replens to be too slippery or lubricating, based on our results, we assume that the lubricating abilities of KY, Pre-Seed and Astroglide would be even more unacceptable, at least for this cohort of users. As noted above, this approach can be applied to novel formulations prior to expensive safety testing.

Descriptive analysis also differs from qualitative work due to the strong emphasis placed on concept formation and alignment among participants. Concept formation is based on abstraction and generalization where an individual learns to extract commonalities among a set of stimuli and applies them to new stimuli [29]. Thus, training serves two roles. First, reference stimuli help a participant decide what is or is not included within a concept. This is important when a concept has a fuzzy boundary (e.g., falls within the grey zone). Second, the consensus step aligns concepts across panelists to ensure they are using the same semantic label for the same underlying percept. For example, the terms ‘snotty’, ‘ropy’ and ‘spinnbarkeit’ all describe the same phenomenon. Likewise, participants were able to rate ‘stickiness’ based on mineral oil and lanolin references that demonstrated ‘not sticky’ and ‘very sticky’, respectively. Participants in previous microbicide acceptability studies have used sticky to describe their experiences with gels [22,23,26,30]. Without concept alignment, assuming sticky means the same things to all populations may not be valid. Thus, findings that stickiness is both a positive and a negative sensation could conceivably result from word usage rather than, or in addition to, actual cultural differences with regard to preferred sensations. Present methods help to better define these characteristics. Additionally, labels depend on context during concept formation, especially if extreme references are not included (e.g., Gynol seems very sticky if one has never rubbed lanolin between the fingers). Current work not only extends previous studies by identifying key descriptors that differentiate among vaginal products, but also provides other references to help align concepts. This approach may have particular utility when conducting cross-cultural studies; work on foods suggests textural attributes are relatively stable across cultures even when different languages are used [31,32].

Finally, in contrast to focus groups, descriptive analysis does not measure a participant’s acceptability of the products. Via training, the participants are turned into calibrated instruments, analogous to a pH meter or rheometer, so asking for acceptability estimates is no longer appropriate [12]. Thus, this approach only provides insight when conducted within a larger testing scheme that includes consumer trials and in vivo focus groups. With both sets of data in hand, it becomes possible to interpret consumer data in light of the descriptive data. For example, if Product A is stickier than Product B in descriptive analysis, and B is

preferred in an acceptance trial, then we can infer that A was not liked because it is too sticky. Critically, this provides more insight than acceptance testing alone, as it provides the formulation scientist with guidance that pursuing formulations stickier than A are not a productive use of resources. Of course, in practice, Products A and B likely differ in more than one attribute. Nonetheless, this approach provides novel information that can be used in formulation optimization.

Cultural differences in sexual and intravaginal practices suggest a single microbicide will not be acceptable in all regions and thus several formulations must be developed. We believe descriptive analysis has the potential to inform the rational design of microbicide gels. Multiple formulations can be evaluated in a less expensive, intrusive and time-consuming manner, and generate quantitative data to model acceptability based on key factors like slipperiness or stickiness. Currently, we are conducting additional work to determine whether our quantitative data correlates with rheological measures (e.g., Refs. [33,34]) to develop models of how changes in formulation alter perception and acceptability.

Our study has some obvious limitations given that these products are meant for use in the vagina, where they will be subjected to dilution, various shear stresses and pH changes [33,35]. For example, implications of attributes such as air bubbles, graininess, ropiness and peaking are hard to interpret since microbicides are used intravaginally and not in the hand. Also, tactile response in the hand/arm may not be an appropriate surrogate for sensation in the vagina. However, designing a study to collect quantitative data on these products in the intended usage area would be very difficult given that only one product can be evaluated within a session and subsequent sessions would have to be several days apart to allow restoration of the vagina to its normal physiological state. Data collected here is not intended to replace studies of these products in the vagina but instead is meant to supplement other work (e.g., Refs. [8,10,20]) by providing quantitative data as a basis for better understanding of human perception of these products. Before this approach can be applied broadly, additional work is needed to confirm whether evaluation on the hand or arm predicts response in the vagina with regard to either sensation or acceptability.

In summary, we identified quantitative differences among vaginal products using a method that had not previously been applied to microbicide formulation or development. As part of a broader repertoire of methods, this approach has the potential to inform future formulation efforts.

## Acknowledgments

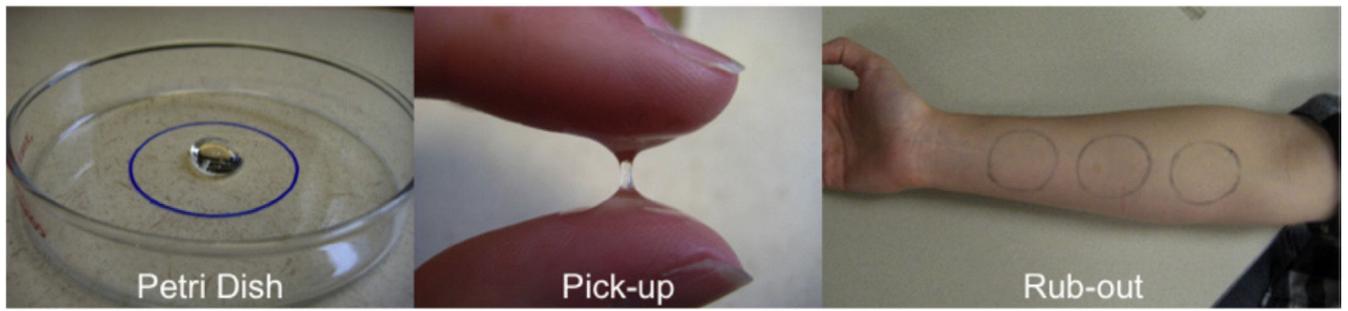
The authors thank the study participants for their time and participation.

## References

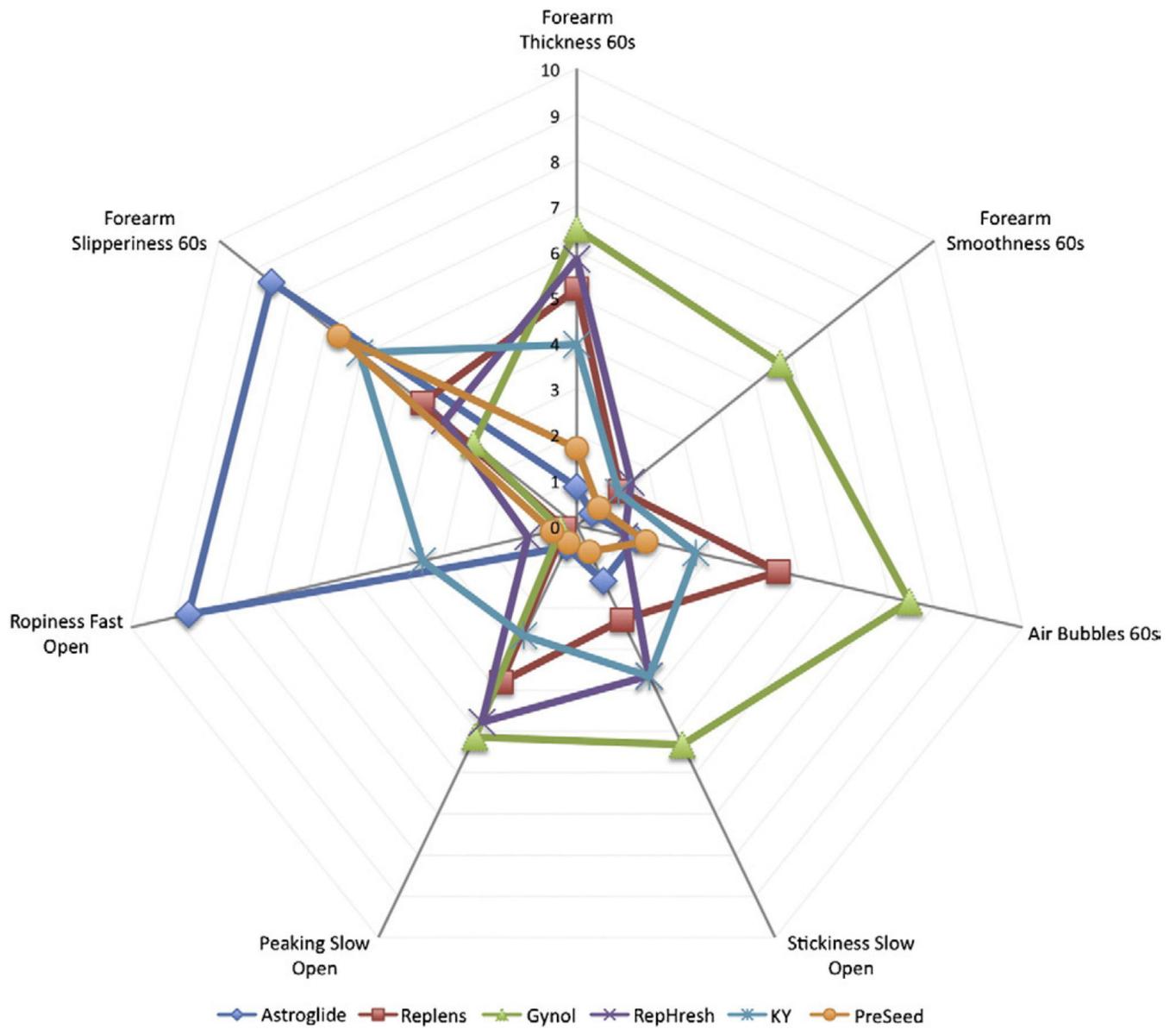
1. Lowndes, CM.; Alary, M.; Belleau, M., et al. West Africa HIV/AIDS epidemiology and response synthesis: implications for prevention. Washington, DC: World Bank; 2008.
2. CDC. [cited 24 June] CDC HIV/AIDS Fact Sheet: HIV/AIDS among Women. [internet]. 2008. Available at:<http://www.cdc.gov/hiv/topics/women/resources/factsheets/pdf/women.pdf>
3. AVAC. Fact Sheet: Microbicides for HIV Prevention. [internet]. 2010 Available at: <http://www.avac.org/ht/a/GetDocumentAction/i/5849>.

4. AVAC. Microbicides clinical trials table as of June 2010. [internet]. 2010 Available at:<http://www.avac.org/ht/a/GetDocumentAction/i/3109>.
5. Abdool Karim Q, Abdool Karim SS, Frohlich JA, et al. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science*. 2010; 329(5996): 1168–1174. [PubMed: 20643915]
6. Stone A. Microbicides: a new approach to preventing HIV and other sexually transmitted infections. *Nat Rev Drug Discov*. 2002; 1:977–985. [PubMed: 12461519]
7. Garg S, Tambwekar KR, Vermani K, et al. Development pharmaceuticals of microbicide formulations: Part II. Formulation, evaluation, and challenges. *Aids Patient Care St*. 2003; 17:377–399.
8. Morrow KM, Ruiz MS. Assessing microbicide acceptability: a comprehensive and integrated approach. *AIDS Behav*. 2008; 12:272–283. [PubMed: 17592763]
9. Hammett TM, Mason TH, Joanis CL, et al. Acceptability of formulations and application methods for vaginal microbicides among drug-involved women: results of product trials in three cities. *Sex Transm Dis*. 2000; 27:119–126. [PubMed: 10676980]
10. Morrow KM, Rosen R, Richter L, et al. The acceptability of an investigational vaginal microbicide, PRO 2000 Gel, among women in a phase I clinical trial. *J Women's Health (Larchmt)*. 2003; 12:655–666. [PubMed: 14583106]
11. Civile GV, Dus Claire A. Evaluating tactile properties of skincare products: a descriptive analysis technique. *Cosmetics & Toiletries*. 1991; 106(5)
12. Lawless, HT.; Heymann, H. *Sensory evaluation of food: principles and practices*. New York: Chapman & Hall; 1998. p. xvii. 819
13. Murray JM, Delahunty CM, Baxter IA. Descriptive sensory analysis: past, present and future. *Food Res Int*. 2001; 34:461–471.
14. Stone, H.; Sidel, J. *Sensory Evaluation Practices*. London: Academic Press; 2004.
15. Schwartz NO. Adaptation of the sensory texture profile method to skin care products. *J Texture Stud*. 1974; 6:33–42.
16. Nguyen D, Lee H, Poast J, et al. Preventing sexual transmission of HIV: anti-HIV bioregulatory and homeostatic components of commercial sexual lubricants. *J Biol Regul Homeost Agents*. 2004; 18:268–274. [PubMed: 15786693]
17. Zubowicz EA, Oakes JK, Short MB, et al. Adolescents' descriptions of the physical characteristics of microbicide surrogates and experiences of use. *J Women's Health (Larchmt)*. 2006; 15:952–961. [PubMed: 17087619]
18. Hootman, RC., editor. *Manual on descriptive analysis testing for sensory evaluation*. Philadelphia, (PA): ASTM; 1992. p. 52
19. Findlay CJ, Castura JC, Lesschaeve I. Feedback calibration: a training method for descriptive panels. *Food Qual Prefer*. 2007; 18:321–328.
20. Hoffman S, Morrow KM, Mantell JE, et al. Covert use, vaginal lubrication, and sexual pleasure: a qualitative study of urban U.S. Women in a vaginal microbicide clinical trial. *Arch Sex Behav*. 2010; 39:748–760. [PubMed: 19636696]
21. Cook RL, Downs JS, Marrazzo J, et al. Preferred characteristics of vaginal microbicides in women with bacterial vaginosis. *J Women's Health (Larchmt)*. 2009; 18:1163–1167. [PubMed: 19630538]
22. Bentley ME, Fullem AM, Tolley EE, et al. Acceptability of a microbicide among women and their partners in a 4-country phase I trial. *Am J Public Health*. 2004; 94:1159–1164. [PubMed: 15226137]
23. Woodsong C, Alleman P. Sexual pleasure, gender power and microbicide acceptability in Zimbabwe and Malawi. *AIDS Educ Prev*. 2008; 20:171–187. [PubMed: 18433322]
24. Wang Y, Liao SS, Weeks MR, et al. Acceptability of hypothetical microbicides among women in sex establishments in rural areas in Southern China. *Sex Transm Dis*. 2008; 35:102–110. [PubMed: 17767093]
25. Mantell JE, Myer L, Carballo-Diequez A, et al. Microbicide acceptability research: current approaches and future directions. *Soc Sci Med*. 2005; 60:319–330. [PubMed: 15522488]

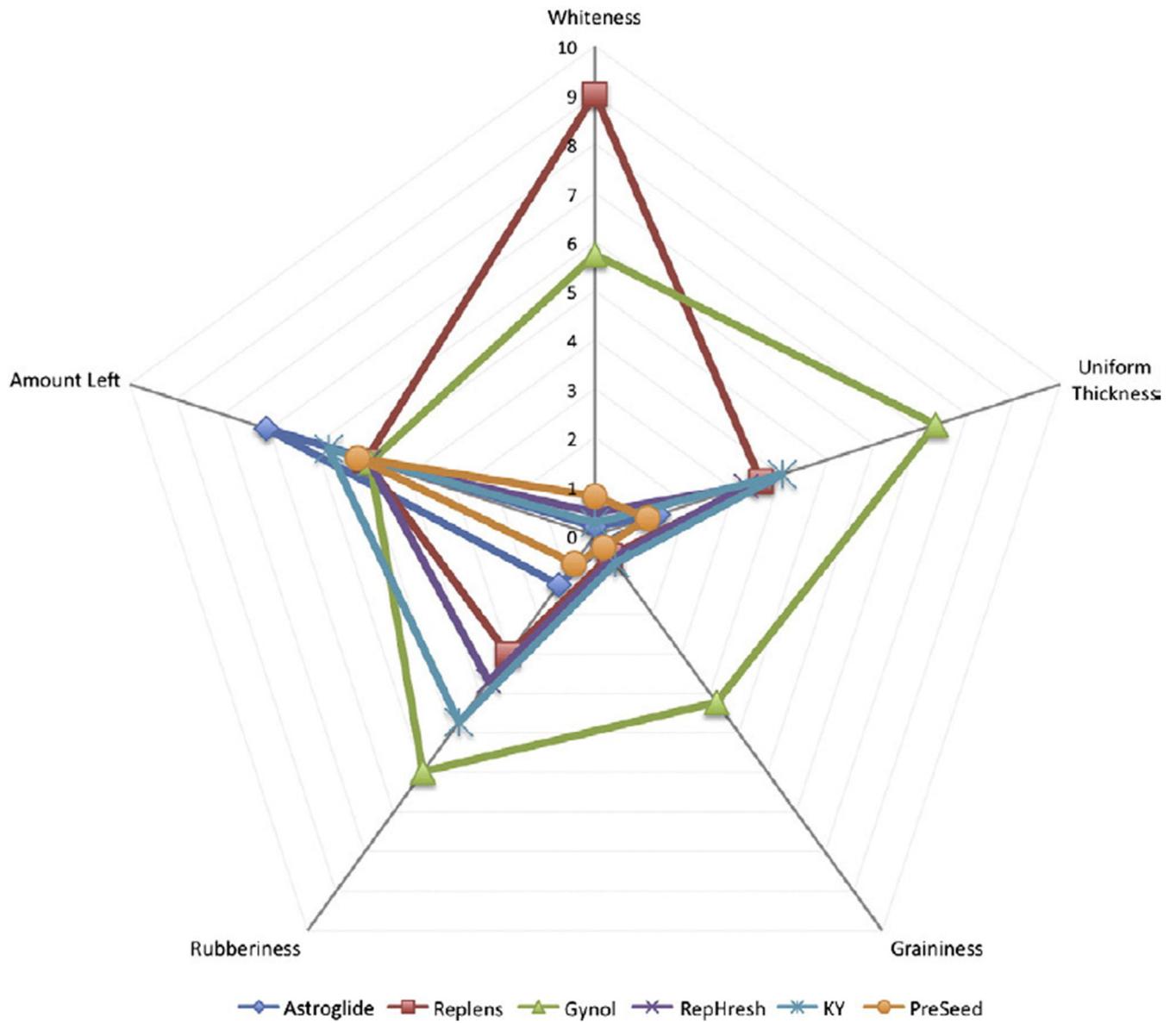
26. Rosen RK, Morrow KM, Carballo-Diequez A, et al. Acceptability of tenofovir gel as a vaginal microbicide among women in a phase I trial: a mixed-methods study. *J Women's Health (Larchmt)*. 2008; 17:383–392. [PubMed: 18328009]
27. Hoffman S, Cooper D, Ramjee G, et al. Microbicide acceptability: insights for future directions from providers and policy makers. *AIDS Educ Prev*. 2008; 20:188–202. [PubMed: 18433323]
28. Braunstein S, van de Wijert J. Preferences and practices related to vaginal lubrication: implications for microbicide acceptability and clinical testing. *J Women's Health (Larchmt)*. 2005; 14:424–433. [PubMed: 15989415]
29. O'Mahony, M. Descriptive analysis and concept alignment. In: Lawless, HT.; Klein, BP., editors. *Sensory science theory and applications in foods*. New York: M. Dekker; 1991. p. 223-267.
30. Bentley ME, Morrow KM, Fullem A, et al. Acceptability of a novel vaginal microbicide during a safety trial among low-risk women. *Fam Plann Perspect*. 2000; 32:184–188. [PubMed: 10942354]
31. Blancher G, Le S, Sieffermann JM, et al. Comparison of visual appearance and texture profiles of jellies in France and Vietnam and validation of attribute transfer between the two countries. *Food Qual Prefer*. 2008; 19:185–196.
32. Tu VP, Valentin D, Husson F, et al. Cultural differences in food description and preference: contrasting Vietnamese and French panellists on soy yogurts. *Food Qual Prefer*. 2010; 21:602–610.
33. Owen DH, Peters JJ, Katz DF. Rheological properties of contraceptive gels. *Contraception*. 2000; 62:321–326. [PubMed: 11239620]
34. Owen DH, Peters JJ, Katz DF. Comparison of the rheological properties of Advantage-S and Replens. *Contraception*. 2001; 64:393–396. [PubMed: 11834239]
35. Owen DH, Peters JJ, Lavine ML, et al. Effect of temperature and pH on contraceptive gel viscosity. *Contraception*. 2003; 67:57–64. [PubMed: 12521660]



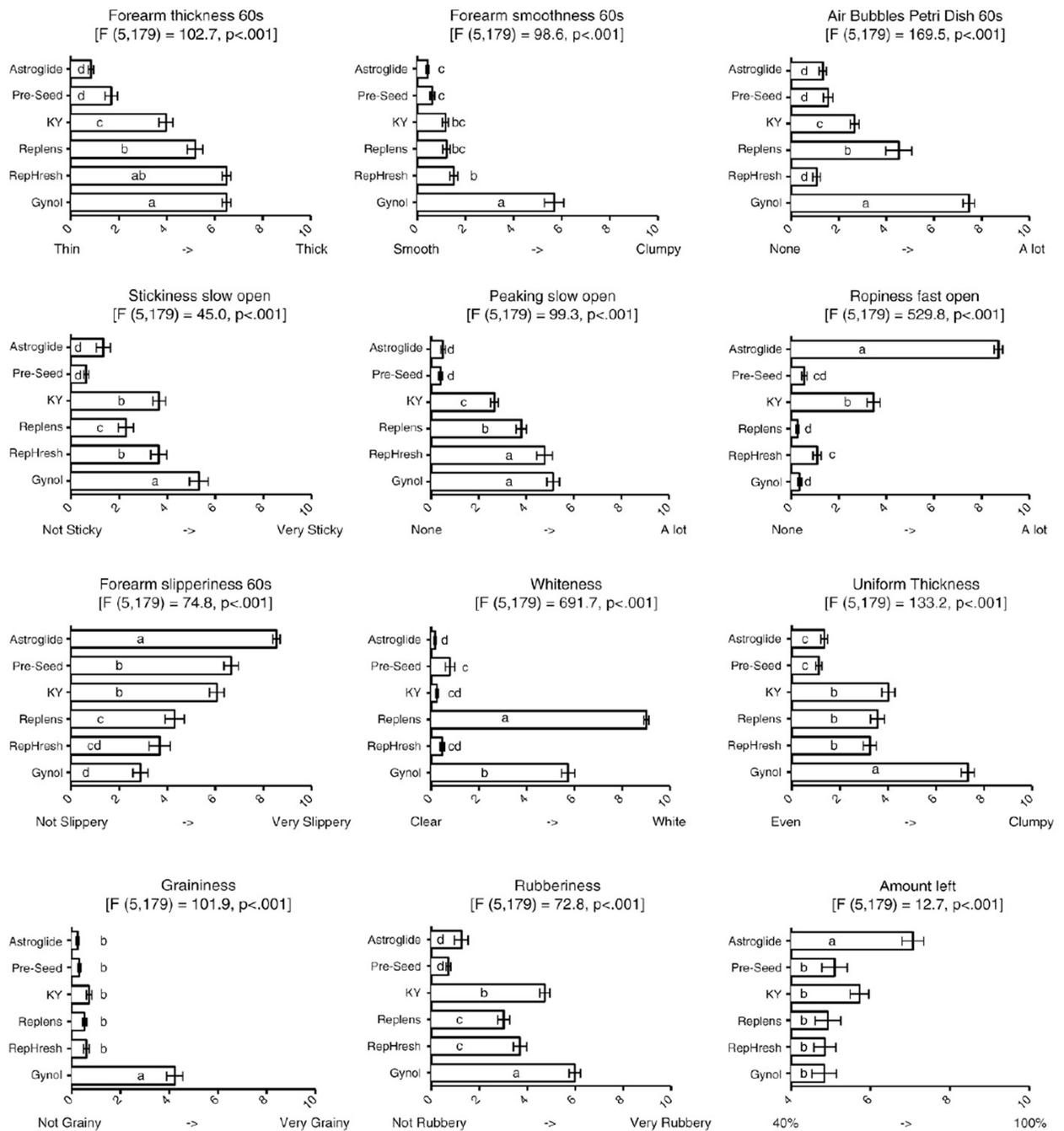
**Fig. 1.**  
Illustration of three evaluation phases.



**Fig. 2.** Differences in major attributes across six OTC vaginal products. See Table 1 for precise definitions of terms and anchors. Statistics are provided in Fig. 4. Please note that ‘smoothness’ is reverse coded (a higher value is more clumpy).



**Fig. 3.** Differences in minor attributes across six OTC vaginal products. See Table 1 for precise definitions of terms and anchors. Statistics are provided in Fig. 4. Please note that ‘uniform thickness’ is reverse coded (a higher value is more clumpy).



**Fig. 4.** Differences in attributes in six OTC vaginal products. Attributes are the same as those shown in Figs. 2 and 3. Within an attribute, products sharing a letter are not significantly different (Tukey's honestly significant difference test, p<.05).

**Table 1**

Descriptive lexicon developed by participants for OTC vaginal products, including consensus definitions and reference standards

<b>Evaluation</b>	<b>Attribute</b>	<b>Low anchor (reference)</b>	<b>High anchor (reference)</b>	<b>Definition</b>
Petri dish				
	White	Clear (mineral oil)	Whiteness (shaving cream)	Initial opacity of the product
	Thickness	Thin (mineral oil)	Thick (lanolin)	Degree to which the product maintained shape
	Peaking	None (mineral oil)	A lot (lanolin)	Ability of and degree to which the product stands up when tapped
	Ropiness	None (mineral oil)	A lot (Probe®)	Ability of the product to string between the finger and remaining product
	Graininess	Not grainy (mineral oil)	Very grainy (apricot scrub)	Degree of hard particulates sensed in the product
	Smoothness	Smooth (mineral oil)	lumpy (tapioca pudding)	Ability of the product to spread evenly on a surface
	Air bubbles	None (mineral oil)	A lot (shaving cream)	Amount of visual bubbles of air in the product
	Uniform thickness	Uniform thickness (mineral oil on petri dish)	Clumpy (glue stick on petri dish)	Visual evaluation of even spread on surface
	Stickiness	Not sticky (mineral oil)	Very sticky (lanolin)	Degree of force necessary to remove finger
Fingers				
	Stickiness	Not sticky (mineral oil)	Very sticky (lanolin)	Degree of force necessary to remove finger
	Peaking	None (mineral oil)	A lot (lanolin)	Ability of and degree to which the product stands up when tapped
	Ropiness	None (mineral oil)	A lot (Probe®)	Ability of the product to string between the finger and remaining product
	Rubberiness	Not rubbery (mineral oil)	Very rubbery (dried rubber cement balls)	Degree of elasticity displayed when compressed
	Thickness	Thin (mineral oil)	Thick (lanolin)	Amount of product layered between fingers
	Smoothness	Smooth (mineral oil)	Clumpy (tapioca pudding)	Evenness of product layered between fingers
	Slipperiness	Not slippery (lanolin)	Very slippery (Gun Oil®)	Degree of resistance experienced during spreading
Forearm				
	Thickness	Thin (mineral oil)	Thick (lanolin)	Amount of product layered between finger and skin
	Smoothness	Smooth (mineral oil)	Clumpy (tapioca pudding)	Evenness of product layered between finger and skin
	Slipperiness	Not slippery (lanolin)	Very slippery (Gun Oil®)	Degree of resistance experienced during spreading
	Air bubbles	None (mineral oil)	A lot (shaving cream)	Amount of visual bubbles of air in the product
	Amount left	0%	100%	Amount of product left on forearm compared to initial application