

## Response to Commentary by Spielmans



Flibanserin was approved by the US FDA in 2015 and by Health Canada in 2018 for the treatment of acquired, generalized hypoactive sexual desire disorder (HSDD) in premenopausal women. Flibanserin was subsequently approved by Health Canada for treating acquired, generalized HSDD in naturally postmenopausal women who are 60 years of age or younger. Yet, data regarding flibanserin's efficacy and safety continue to be misunderstood. Our recent post-hoc analyses<sup>1</sup> of the effects of flibanserin further characterized changes in aspects of sexual function beyond sexual desire. Spielmans' commentary<sup>2</sup> on our publication attempts to characterize flibanserin's efficacy as being "unimpressive" and "likely not clinically meaningful". Spielmans opines that the reported effect sizes in our analyses were too small, that our discussion comparing effect sizes and NNT estimates for flibanserin to those of antidepressants is irrelevant, that the PGI-I data have no clinical relevance, and that the Female Sexual Function Index (FSFI) has questionable validity as a measure of sexual function in women with HSDD.

Unfortunately, these assertions are based upon erroneous assumptions that expose a lack of understanding of the condition of HSDD and are inconsistent with the intended use of the statistical metrics that Spielmans vaunts as proof of flibanserin's poor efficacy. In each of his arguments about quantitative changes, Spielmans ignores the substantial impact of HSDD and the effects of improving sexual desire and symptoms of distress associated with this condition. Previous studies have shown that women who report symptoms of low sexual desire and concomitant distress experience relationship difficulties with their partner, have negative perceptions of body image, and decreased self-confidence with feelings of frustration, loss, and anxiety.<sup>3,4</sup> In fact, HSDD has been associated with decreased health-related quality of life that is comparable to that of diabetes or back pain.<sup>5</sup> In the context of such chronic medical conditions that substantially decrease health-related quality of life, numerically small improvements can result in clinically meaningful benefit.

Statistical analyses by themselves cannot determine the meaningfulness of a change to a given patient. This is the reason for using an outcome like the Patient Global Impression of Improvement (PGI-I) so that patients can report firsthand whether they perceived any therapeutic benefit. Spielmans criticizes the use of this validated patient-reported outcome because overall benefit assessments included patients who reported "minimally improved" or better. Spielmans contends that "nobody seeks treatment aiming for minimal improvement" and self-defines "clinically meaningful" improvement as PGI-I responses of "much improved" or "very much improved." As clinicians who care for patients with HSDD and researchers who investigate the etiologies, effects, and treatments related to this condition, we

would never impose an external measure that dictates the goals of a patient's therapy or defines "success" versus "failure." The consensus process of care for the management of HSDD in women developed by the International Society for the Study of Women's Sexual Health emphasizes the importance of patient and partner preferences and goals.<sup>6</sup>

With regard to Spielmans' critique of the PGI-I results, a previous publication by Simon et al.<sup>7</sup> clearly demonstrates that a greater percentage of patients in the flibanserin group consistently reported that their HSDD was "much" or "very much" improved when compared to the placebo group. Further, fewer patients in the flibanserin group reported that their condition did not change or became worse when compared to the placebo group. The consistency of the benefit differential between flibanserin and placebo at any "cut point" on the PGI-I scale demonstrates a robust, clinically meaningful effect. It would be more concerning if the benefit of flibanserin was only apparent in those reporting minimal improvement, but this is not the case.

In analyzing the FSFI data, it is not surprising that the magnitude of the standardized effect size calculations (Cohen's *d*) was greatest for the sexual desire domain, since flibanserin is intended to treat the condition of HSDD and patients were specifically diagnosed with HSDD. The standardized effect size for the arousal domain was comparable to that of the desire domain, since these aspects of sexual function are more closely linked and patients who reported symptoms of sexual arousal disorder were not excluded from the study as long as they identified their distressing low sexual desire as being more severe. Women with orgasmic and/or pain disorders were excluded from the study. Thus, there was no expectation of benefit for these domains. Nevertheless, patients treated with flibanserin still reported a benefit for orgasm that was highly statistically significant in terms of mean domain scores between treatment and placebo groups, as well as standardized effect sizes.

Spielmans' characterization of the calculated standardized effect sizes being "small" is also based on an overly simplistic reliance on Cohen's "rule of thumb." As we have already noted in our discussion, "...Cohen's benchmarks for small (0.2), medium (0.5) and large (0.8) effect sizes are arbitrary and should only be used if no other indices of standardized effect size are available."<sup>8</sup> Further, as we also point out in the discussion, "Evaluation of subjective outcomes such as sexual desire can also be associated with lower effect size estimates compared to objective indices like blood pressure, serum cholesterol, or glucose levels."<sup>8,9</sup> As emphasized by Leucht and colleagues, "The increment of improvement by drug over placebo must be viewed in the context of the disease's seriousness, suffering induced, natural course, duration, outcomes, adverse events and societal values."<sup>9</sup> This is

precisely the reason that we provided context to our analyses. While flibanserin has a different mechanism of action compared to other currently approved psychotropic drugs, they are all centrally acting medications that treat multifactorial conditions that also commonly use subjective assessments. Thus, we believe that our comparisons are valid and clinically relevant.

Spielmann argues that the standardized effect sizes for flibanserin are numerically smaller than the standardized effect sizes of antidepressants and medications for anxiety and obsessive-compulsive disorder, all of which we disclose in our discussion. Yet, he readily concedes that a given effect size may mean different things when considered in the context of HSDD treatments versus depression treatments. Given this concession, Spielmann's implication that effect sizes for flibanserin are too small to be meaningful is inconsistent and baseless; one cannot assert inferiority of efficacy based on numerically smaller effect sizes while also asserting that effect sizes should be interpreted in the context of the condition being treated. For reasons already stated, we would agree with Spielmann on the latter point. Our intent in providing the comparison to other commonly used psychotropic medications was to merely point out that the standardized effect sizes for these types of drugs are in the small to low-medium range. Thus, expectations of numerically larger effect sizes may be unrealistic.

Similarly, although Spielmann admits that the NNT values for flibanserin compare favorably to those of antidepressants, he is dismissive of the reported NNT values, asserting that the comparison to the NNT values of antidepressants is not relevant. Notably, Spielmann does not state that flibanserin is superior or even comparable to antidepressants based on NNT even though this would be a valid conclusion if one were to apply his line of argumentation that the numerical values of effect sizes are a direct reflection of efficacy or clinical relevance. As stated in the previous paragraph, we provided the comparison to antidepressants merely to provide perspective. As noted by Spielmann, calculations for NNT were done for completers rather than the full analysis set. This was due to the fact that the primary reviewer of our manuscript was very concerned about the impact of early discontinuation. Including only study completers likely enriched the treatment cohort with those experiencing benefit. This further emphasizes the need to interpret any measure of effect size in context and not as an absolute measure of clinical efficacy or relevance. Importantly, we disclosed the assumptions and parameters of our analyses and discussed the limitations of our findings to allow the reader to formulate their own interpretation and conclusions.

Next, Spielmann asserts that aside from sexual desire, other domains of the FSFI are invalid for women who are sexually inactive. He acknowledges that participants in the flibanserin trials were required to engage in sexual activity on a monthly basis but expresses doubt that everyone was compliant in this regard due to their HSDD. In reality, premenopausal patients at trial entry reported an average of 2.7 satisfying sexual events per month,

while postmenopausal women reported an average of 2.0 such events per month. Clinicians and researchers who are experienced in sexual health know very well that women continue to engage in sexual activity for a myriad of reasons that do not necessarily correlate with their level of sexual desire.<sup>10,11</sup> It should also be made clear that those who participated in the flibanserin trials were in "a stable, monogamous, heterosexual relationship that was secure and communicative, for at least 1 year prior to the Screen Visit" and that "the relationship had to be with the same partner who was sexually functional, both psychologically and physically, and the partner was expected to be physically present (i.e., available for sexual activity at some time during a 24-hour day) at least 50% of each month during the four-week Screen period and the 24-week efficacy period of the trial." Thus, the participants in the flibanserin trials were in supportive, non-coercive, stable relationships with a sexually functional partner and already engaged in sexual activity at least once per month prior to entering the trials.

Spielmann further expresses doubt about the content validity of the FSFI based upon a previous publication by Revicki et al.<sup>12</sup> Extrapolating from these published data, Spielmann determined that 33 of 75 (44%) women with HSDD reported that the questions in the desire domain did not entirely capture their sexual desire and/or interest problems. Spielmann opines that this is "not impressive." Unfortunately, like much of Spielmann's argumentation throughout his commentary, this line of reasoning is specious. For any patient-reported outcome, content validity is determined by the extent to which the instrument measures the concepts that are most significant and relevant to a patient's condition. In the Revicki study, questions 1 and 2 of the FSFI desire domain were endorsed as being relevant by 100% and 93% of the study cohort, respectively. A psychometric instrument need not "entirely capture" every aspect of what the patient is experiencing. Thus, lack of completeness does not necessarily equate with lack of accuracy. For example, diagnostic tests do not completely and comprehensively capture every aspect of a person's health, but this does not invalidate the diagnostic test and its utility in making clinically relevant observations.

Spielmann also asserts that the FSFI was originally developed to assess women with sexual arousal disorder and has not been well studied in women with HSDD. While it is true that Rosen et al. initially validated the FSFI in 2000 by comparing a cohort of women without sexual dysfunction against a cohort of women with sexual arousal disorder, it is well known that the FSFI was subsequently validated in women with HSDD and also in women with orgasmic disorder just 3 years later.<sup>13,14</sup> The most important aspect that makes a psychometric instrument clinically useful in drug trials is its ability to measure change in sexual function that can be correlated to severity of a given condition. Indeed, each domain of the FSFI has been shown to be a sensitive measure of sexual dysfunction with statistically significant decreases in domain scores for women with orgasmic disorder or HSDD compared to age-matched controls.<sup>14</sup> While there are

certainly other patient-reported outcomes that have been validated for the assessment of female sexual function, the FSFI has become the most widely used instrument in this context, appearing in over 1,000 publications in at least 20 different translations to date.<sup>15</sup>

In the latter part of his commentary, Spielmans expresses concern that the difference in the cumulative rates of somnolence, fatigue and sedation is “notable” (21% for flibanserin vs. 8% for placebo) and that the possibility of severe hypotension and syncope occurring with concomitant alcohol use is “problematic” given that a “substantial percentage of women. . . consume alcohol at least occasionally.” Spielmans ignores the fact that flibanserin is intended to be dosed at bedtime so that sedation-related adverse events do not become problematic. Even with flibanserin’s extended half-life of 11 hours, a dedicated study found that there was no impairment of cognitive function or driving performance the following morning in subjects who were administered the therapeutic dose of flibanserin at bedtime.<sup>16</sup> With regard to the alcohol warning, flibanserin is not unique. Whether or not there is a specific warning in the package insert, concomitant alcohol use is not advised with any psychotropic medication, medications that lower blood pressure, treat diabetes, or even erectile dysfunction due to concerns over adverse events that include CNS depression, hypotension, or hypoglycemia. Flibanserin’s overall safety profile has been extensively studied.<sup>17</sup> In particular, the alcohol interaction is extremely well characterized in 4 separate studies such that a minimum time interval between alcohol use and safe flibanserin dosing has been defined.<sup>17,18</sup>

We wholeheartedly agree with Spielmans that measures of relationship satisfaction and overall well-being should be examined as part of the overall benefit assessment for therapies targeting sexual dysfunctions. We also agree that more and better treatments and assessment tools are needed and would welcome financial resources from both public and private funding organizations, as well as pharmaceutical and biotech companies that could assist with this endeavor. However, for the reasons stated above, we strongly disagree with the interpretation that flibanserin’s efficacy data is “underwhelming”. With specific relevance to our recent publication on the FSFI domain data, it is important to emphasize the holistic perspective that individual domains of sexual function are interdependent and represent a spectrum of sexual response that can be measurably improved if distressing low sexual desire is ameliorated.

In closing, we support dialogue and debate of research findings and issues related to healthcare. Medical journals are important platforms and venues for such exchanges. However, dialogue and debate must be well-informed, and participants must exercise integrity in order to be edifying and worthwhile. Statistical analyses and associated findings should be used as a tool to aid interpretation and provide context rather than a bludgeon that only seeks to spare or destroy based upon some predetermined cut-off. Although Leucht and colleagues were thoughtful enough to include societal values as an integral part of

a drug’s benefit-risk assessment, this aspect is often overlooked. The validity of sexual dysfunctions in women have been openly questioned for quite some time. In addition, it is our experience that women have often been mischaracterized as being less capable of making appropriate choices in their medical care, providing an accurate history, evaluating benefits and risks of various therapies, giving informed consent for treatment, and accurately communicating efficacy and adverse events. To counteract these deeply entrenched biases, there is a critical need for education in our field for both the general public and for researchers and medical care providers. We invite the editorial board of the SMOA to actively engage in upholding the highest standards of professionalism to further the dissemination of accurate information and well-informed discussion.

Respectfully,

James A. Simon, MD

Anita H. Clayton, MD

Irwin Goldstein, MD

Sheryl A. Kingsberg, PhD

Marla Shapiro, MDCM

Sejal Patel, PharmD

Noel N. Kim, PhD

James A. Simon, MD,<sup>1</sup> Anita H. Clayton, MD,<sup>2</sup> Irwin Goldstein, MD,<sup>3</sup> Sheryl A. Kingsberg, PhD,<sup>4</sup> Marla Shapiro, MDCM,<sup>5</sup> Sejal Patel, PharmD,<sup>6</sup> and Noel N. Kim, PhD<sup>7</sup>

<sup>1</sup>IntimMedicine Specialists and George Washington University School of Medicine, Washington, DC, USA

<sup>2</sup>University of Virginia School of Medicine, Charlottesville, VA, USA

<sup>3</sup>San Diego Sexual Medicine and Alvarado Hospital, San Diego, CA, USA

<sup>4</sup>University Hospitals Cleveland Medical Center and Case Western Reserve University, Cleveland, OH, USA

<sup>5</sup>University of Toronto, ON, Canada

<sup>6</sup>Sprout Pharmaceuticals, Inc., Raleigh, NC, USA

<sup>7</sup>Institute for Sexual Medicine, San Diego, CA, USA

**Corresponding Author:** Noel N. Kim, PhD, Institute for Sexual Medicine, 6330 Nancy Ridge Drive, Suite 105, San Diego, CA 92121, USA; E-mail: [nkim@ismlab.org](mailto:nkim@ismlab.org)

<https://doi.org/10.1016/j.esxm.2022.100585>

## REFERENCES

1. Simon JA, Clayton AH, Goldstein I, et al. Effects of flibanserin on subdomain scores of the Female Sexual Function Index in women with hypoactive sexual desire disorder. *Sex Med* 2022. doi: [10.1016/j.esxm.2022.100570](https://doi.org/10.1016/j.esxm.2022.100570). E-pub ahead of print.

2. Spielmans GI. Pooled analysis confirms flibanserin's unimpressive efficacy, raises measurement questions. *Sex Med* 2022 Press. doi: [10.1016/j.esxm.2022.100579](https://doi.org/10.1016/j.esxm.2022.100579).
3. Kingsberg SA. Attitudinal survey of women living with low sexual desire. *J Womens Health (Larchmt)* 2014;23:817–823. doi: [10.1089/jwh.2014.4743](https://doi.org/10.1089/jwh.2014.4743).
4. Simon JA, Athavale A, Ravindranath R, et al. Assessing the burden of illness associated with acquired generalized hypoactive sexual desire disorder. *J Womens Health (Larchmt)* 2022;31:715–725. doi: [10.1089/jwh.2021.0255](https://doi.org/10.1089/jwh.2021.0255).
5. Biddle AK, West SL, D'Aluisio AA, et al. Hypoactive sexual desire disorder in postmenopausal women: quality of life and health burden. *Value Health* 2009;12:763–772. doi: [10.1111/j.1524-4733.2008.00483.x](https://doi.org/10.1111/j.1524-4733.2008.00483.x).
6. Clayton AH, Goldstein I, Kim NN, et al. The International Society for the Study of Women's Sexual Health process of care for management of hypoactive sexual desire disorder in women. *Mayo Clin Proc* 2018;93:467–487. doi: [10.1016/j.mayocp.2017.11.002](https://doi.org/10.1016/j.mayocp.2017.11.002).
7. Simon JA, Clayton AH, Kim NN, et al. Clinically meaningful benefit in women with hypoactive sexual desire disorder treated with flibanserin. *Sex Med* 2022;10:100476. doi: [10.1016/j.esxm.2021.100476](https://doi.org/10.1016/j.esxm.2021.100476).
8. Bakker A, Cai J, English L, et al. Beyond small, medium, or large: Points of consideration when interpreting effect sizes. *Educ Stud Math*. 2019;102:1–8.
9. Leucht S, Hierl S, Kissling W, et al. Putting the efficacy of psychiatric and general medicine medication into perspective: review of meta-analyses. *Br J Psychiatry* 2012;200:97–106. doi: [10.1192/bjp.bp.111.096594](https://doi.org/10.1192/bjp.bp.111.096594).
10. Meston CM, Buss DM. Why humans have sex. *Arch Sex Behav* 2007;36:477–507. doi: [10.1007/s10508-007-9175-2](https://doi.org/10.1007/s10508-007-9175-2).
11. Meston CM, Hamilton LD, Harte CB. Sexual motivation in women as a function of age. *J Sex Med* 2009;6:3305–3319. doi: [10.1111/j.1743-6109.2009.01489.x](https://doi.org/10.1111/j.1743-6109.2009.01489.x).
12. Revicki DA, Margolis MK, Bush EN, et al. Content validity of the Female Sexual Function Index (FSFI) in pre- and postmenopausal women with hypoactive sexual desire disorder. *J Sex Med* 2011;8:2237–2245. doi: [10.1111/j.1743-6109.2011.02312.x](https://doi.org/10.1111/j.1743-6109.2011.02312.x).
13. Rosen R, Brown C, Heiman J, et al. The Female Sexual Function Index (FSFI): a multidimensional self-report instrument for the assessment of female sexual function. *J Sex Marital Ther* 2000;26:191–208. doi: [10.1080/009262300278597](https://doi.org/10.1080/009262300278597).
14. Meston CM. Validation of the Female Sexual Function Index (FSFI) in women with female orgasmic disorder and in women with hypoactive sexual desire disorder. *J Sex Marital Ther* 2003;29:39–46. doi: [10.1080/713847100](https://doi.org/10.1080/713847100).
15. Meston CM, Freihart BK, Handy AB, et al. Scoring and Interpretation of the FSFI: what can be learned from 20 years of use? *J Sex Med* 2020;17:17–25. doi: [10.1016/j.jsxm.2019.10.007](https://doi.org/10.1016/j.jsxm.2019.10.007).
16. Kay GG, Hochadel T, Sicard E, et al. Next-day residual effects of flibanserin on simulated driving performance in premenopausal women. *Hum Psychopharmacol* 2017;32. doi: [10.1002/hup.2603](https://doi.org/10.1002/hup.2603).
17. Clayton AH, Brown L, Kim NN. Evaluation of safety for flibanserin. *Expert Opin Drug Saf* 2020;19:1–8. doi: [10.1080/14740338.2020.1707804](https://doi.org/10.1080/14740338.2020.1707804).
18. Addyi (flibanserin) tablets, for oral use. Raleigh, NC: Sprout Pharmaceuticals; 2021.