THE JOURNAL OF SEX RESEARCH
https://doi.org/10.1080/00224499.2021.1885601

Re-Analyzing Phase III Bremelanotide Trials for "Hypoactive Sexual Desire Disorder" in Women

Glen I. Spielmans

Department of Psychology, Metropolitan State University

ABSTRACT

Kingsberg et al. described results from two 24-week Phase III trials of bremelanotide for treating hypoactive sexual desire disorder (HSDD) in women. 72.72% of protocol-listed outcomes were not reported by Kingsberg et al., who provided results of 15 secondary measures which were not listed in the study protocols. None of their efficacy outcomes were reported in line with CONSORT data reporting standards and no secondary outcome had a stated rationale or cited evidence of validity. My meta-analysis of the trials' data, based on the FDA New Drug Application, found similar results to Kingsberg et al. However, Kingsberg et al. did not report that a) adverse event-induced study discontinuation was substantially higher on bremelanotide: OR = 11.98, 95% CI = 3.74–38.37, NNH: 6 or b) participants preferred placebo, measured by the combination of both 1) completing a clinical trial and 2) electing to participate in the follow-up open-label study (OR = 0.30, 95% CI = .24-.38, NNH: 4). Bremelanotide's modest benefits on incompletely reported post-hoc measures of questionable validity in combination with participants substantially preferring to take placebo suggest that the drug is generally not useful. Kingsberg et al.'s data reporting and measurement practices were incomplete and lacked transparency.

The fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) was released in 1994 (American Psychiatric Association, 1994). In the late 1990s, when pharmacological treatments to enhance female sexual desire and arousal were in development, the DSM-IV contained the list of "sexual dysfunctions" which could be targeted by such treatments, of which hypoactive sexual desire disorder (HSDD) and female sexual arousal disorder (FSAD) were the most relevant. Drug firms funded the development of measurements for the severity of such "sexual dysfunctions" so that the success of their products could be gauged (Moynihan, 2003). In the DSM-5, published in 2013, HSDD and FSAD were both removed (American Psychiatric Association, 2013). They were replaced by a combined condition of female sexual interest/arousal disorder (FSIAD), a disorder including reduced sexual desire, lack of response to sexual stimuli, and lack of pleasure during sexual activity, impacting at least 75% of sexual encounters and causing significant personal distress over a period of at least six months.

Flibanserin was developed to treat HSDD when the DSM-IV definition was in effect, and became the first drug to receive Food and Drug Administration (FDA) approval to treat HSDD in August 2015. During HSDD's time in the DSM-IV, bremelanotide was also in development. It was approved by the Food and Drug Administration to treat HSDD in June 2019. Thus, there are now two relatively recently approved drugs for HSDD, a condition that no longer exists in the DSM-5. HSDD is still present in the International Classification of Diseases (11th edition), in which it can be applied to either men or women.

A systematic review of flibanserin found evidence of quite modest treatment efficacy versus placebo in terms of relevant rating scale scores and number of monthly satisfying sexual events (Jaspers et al., 2016). Two Phase III placebo-controlled trials formed the final basis of the FDA's approval of bremelanotide in June 2019. There has been no independent analysis of these trials, which is potentially problematic given shortcomings in transparency, reproducibility, and data reporting observed in many scientific fields.

Reproducibility Crisis and Questionable Research Practices

It has become increasingly clear that psychological science often generates published results that other researchers cannot replicate. In perhaps the best-known illustration of this problem, an attempt to replicate 100 studies published in psychology journals resulted in an average reduction of effect size of over 50%. Further, 97% of the original studies yielded statistically significant results; this was true in only 36% of the attempted replications (Open Science Collaboration, 2015). Researchers often engage in questionable research practices or "researcher degrees of freedom" that maximize the odds of finding statistically significant results regarding variables of interest (John et al., 2012; Simmons et al., 2011). One such practice is "data peeking", in which researchers perform statistical analyses at various points of data collection, stopping once they have obtained a statistically significant result. Further, sometimes researchers report data for only a subset of variables. Researchers sometimes change the *a priori* "primary outcome" to a secondary outcome if it fails to achieve statistical

CONTACT Glen I. Spielmans glen.spielmans@metrostate.edu Department of Psychology, Metropolitan State University, 1450 Energy Park Drive, Saint Paul, MN 55108.

significance, and switch a statistically significant secondary outcome to the primary outcome. These practices then lead to focusing on the "interesting" statistically significant results while overlooking data from variables which did not yield significant results (Bradley et al., 2017; Mathieu et al., 2009). These practices are entwined with "HARKing" (hypothesizing after the results are known), wherein a researcher who knows a study's results subsequently tailors the research hypotheses to fit these results (Kerr, 1998). When HARKing, *a priori* hypotheses are silently discarded, leaving readers and researchers unaware of their lack of empirical support, impeding scientific progress. Further, HARKing is circular reasoning, as one examines the data to generate hypotheses *post hoc*, then claims the hypotheses are supported by the data that generated the hypotheses. Post hoc analyses may generate interesting new leads, but this is not the same as confirming a hypothesis made *before* data were collected.

A survey of 2,155 academic psychologists in the USA inquired about engagement in 10 questionable research practices (QRPs). Over 60% of respondents indicated they had not reported all dependent variables in a paper, over half reported that they had stopped data collection upon learning their results were statistically significant, and nearly half admitted to selectively reporting studies that generated statistically significant results while not reporting studies that lacked statistical significance (John et al., 2012). Respondents reported that other researchers were more likely to engage in several of these practices than themselves. As a whole, John et al.'s results suggest that QRPs occur frequently.

Data analysis offers many opportunities to generate statistically significant results. One can control for any number of covariates (e.g., gender, age, initial symptom severity, etc.), perform interim data analyses as data are collected (then stop when a significant result is obtained), and utilize any number of dependent variables (Simmons et al., 2011). Each of these procedures raises the risk of a type I error (a "false positive"), in which the null hypothesis is rejected although it is actually true. Researchers should provide transparency of measurement, clearly describing all measures and providing evidence for their validity. It should also be clear which measures were *a priori* and which were post hoc. Steps that minimize transparency of measurement are questionable measurement practices (QMPs) (Flake & Fried, in press). As stated by Flake and Fried, "A lack of information about the measures in a study introduces uncertainty in all aspects of a study's validity (p. 8)."

On a related note, data from continuous rating scales are sometimes transformed into binary outcomes such as "treatment response." Such binary outcomes make the most sense when the underlying construct is truly yes/no (e.g., alive/dead, pregnant/not pregnant). Continuous rating scales are validated using reliability and validity assessments based on their use as continuous measures, not on the measurement properties of various ways in which the scale is dichotomized (MacCallum et al., 2002). Consider a measure of "treatment response" defined as improvement of 50% or more on a continuous rating scale of depressive symptoms. Unless there is good evidence demonstrating that improvement of 50% is meaningfully different than improvement of, say, 45%, then this particular definition of treatment response is arbitrary and very likely less informative than the overall score. At the very least, studies that use dichotomized data based on continuous scale scores should also report the results of the continuous scale as well as any validity data regarding the dichotomized outcome. As stated by MacCallum et al. (2002): "Claims of the existence of types [such as responder/non-responder], and corresponding dichotomization of quantitative scales and analysis of group differences, simply must be supported by compelling results from taxometric analyses" (p. 38).

The *Journal of Sex Research* requires researchers to disclose researcher degrees of freedom that allow flexibility in statistical analyses and thus inflate the risk of type I error (Sakaluk & Graham, 2018). The uptake of such standards varies greatly among journals. Based on the well-documented problems with replicability in psychological research results, transparent reporting of researcher flexibility in handling data analyses is clearly warranted. Problems in replicability are not limited to psychology, with demonstrated replicability problems existing in other fields, including psychiatric genetics (Border et al., 2019), psychiatric gene x environment interaction research (Duncan & Keller, 2011), structural brain-behavior associations (Masouleh et al., 2019), cognitive neuroscience (Szucs et al., 2017), and economics (Camerer et al., 2016).

To combat these problems, study protocols can be preregistered in an online database. Then a peer reviewer or journal editor can check a manuscript under review to see if its measures, methods and proposed statistical analyses align with the study protocol. Kaplan et al. (2015) examined whether study preregistration related to reported study outcomes among clinical trials funded by the National Heart Lung, and Blood Institute (NHLBI). All large NHLBI trials were required to preregister their protocols online. Studies whose results were reported prior to 2000, when preregistration became required, had a 57% chance of finding significant benefit on the primary outcome. After preregistration became mandatory, the rate of positive outcomes on the primary outcome plummeted to 8%. Many of the post-2000 studies had secondary outcomes on which statistically significant benefit was observed. Perhaps the preregistration of primary outcomes prevented some post-hoc switching of primary and secondary outcomes. Unfortunately, changes in study methods, measures, or statistical analyses are often not noticed in peer review (Mathieu et al., 2013). But with publicly available protocols, interested readers can identify these issues after an article is published. Further, results can be published in online databases, regardless of whether the study is published in a journal.

Industry-Funded Trials and CONSORT

Pharmaceutical industry-funded clinical trials have demonstrated several data reporting biases. Overstatement of efficacy via such methods as selective outcome reporting, improperly including ineligible participants or excluding eligible participants in statistical analyses, and using post-hoc data analyses to boost the apparent efficacy of a product are all well-documented problems (Jureidini et al., 2016; McHenry & Amsterdam, 2019; Le Noury et al., 2015; Roest et al., 2015; Spielmans et al., 2013; Spielmans & Parry, 2010; Turner et al., 2008). Discrepancies often exist between clinical trial protocols

and reported results, with measures and statistical analyses added or subtracted post-hoc, frequently leading to inflated efficacy reporting in journal articles (Chan et al., 2004; Mathieu et al., 2009). Further, reporting of adverse events is often inadequate and incomplete (Hughes et al., 2014; Mayo-Wilson et al., 2019b, 2019a). In line with the aforementioned problems, clinical trial reports in journal articles often report greater treatment effects and less risk than data reported to regulatory agencies (Hart et al., 2012; Healy & Cattell, 2003). Thus, incorporating data from regulatory agencies such as the FDA alongside data published in journals often conveys a more comprehensive, likely less biased view of treatment efficacy and efficacy.

Various standards of reporting study participants, methods and outcomes exist, with the CONSORT guidelines often recommended as a good reporting checklist for clinical trials (Schulz et al., 2010). According to CONSORT's website, over half of the core medical journals listed in the Abridged Index Medicus on Pubmed endorse CONSORT reporting guidelines (CONSORT, 2020). CONSORT standards call for publication of summary statistics, effect size, and confidence intervals for all prespecified outcomes; any changes in outcome measures made after protocol submission require a clear explanation. CONSORT also calls for the use of previously validated measures in clinical trials whenever possible (Moher et al., 2010).

Academic authors not directly employed by the drug industry appear in authorship lines of nearly all industry-sponsored clinical trials. This lends the appearance of independent oversight of both the trial and related manuscripts. However, the sponsor typically exercises great influence (or total control) over what statistical analyses are conducted; the sponsor has access to raw data that external authors typically lack (Sismondo & Nicholson, 2009). Also, "independent" authors typically have financial conflicts of interest (COI), such as receiving consulting fees from or owning stock in the sponsor of the trial. There is little reason to believe that the presence of non-corporate authors on industry-sponsored trials improves the transparency or accuracy of data reporting (Jureidini & McHenry, 2020; Matheson, 2016b; Sismondo & Nicholson, 2009). Authors with COIs are certainly not incentivized to cast doubt upon the efficacy and safety of products produced by companies who pay them (Fava, 2016). The mere presence of author COIs does not necessarily imply anything nefarious but is worth noting when reading a clinical trial.

Pharmaceutical firms disseminate research findings in a strategic manner via thoughtfully-designed publication plans that target specific audiences with messages of drug efficacy and safety. Drug firms shepherd the creation of manuscripts by hiring medical writers to create publication-read papers in a timely and marketing-friendly manner (Armstrong, 2006; Jureidini & McHenry, 2020; Matheson, 2016b; Sismondo & Nicholson, 2009). In journal articles, the presence of a medical writer is often denoted with a footnote indicating "editorial support" or a similar term. Internal drug industry documents and accounts from former medical writers note that "editorial support" often involves writing the first draft of the paper before it is passed along to the "authors" (Fugh-Berman, 2010; Logdberg, 2011; Matheson, 2016b; Ross et al., 2008). This raises concerns over the degree to which the listed paper authors can vouch for the underlying data and whether they were analyzed appropriately.

In the spirit of open science and assessing the accuracy and completeness of clinical trial reporting, I examined the extent to which data from the two bremelanotide trials reported in Kingsberg et al. (2019) aligned with a) the *a priori* statistical analyses for efficacy outcomes listed in the clinicaltrials.gov study protocols (ClinicalTrials.gov, 2018a, 2018b) and b) efficacy and dropout results reported in the FDA New Drug Application (NDA) (United States Food and Drug Administration, 2019). Given frequently reported problems with data transparency and incompleteness of reported outcomes in both a) peer-reviewed journal articles in general and b) industry-funded clinical trials in particular, I expected that the published journal article reporting clinical trial results (Kingsberg et al., 2019) would overstate bremelanotide's efficacy to some uncertain extent when compared to the data reported in the NDA. I also expected some uncertain amount of deviation in data reporting between the clinicaltrials.gov protocols and the Kingsberg et al. (2019) paper. Further, I examined the extent to which Kingsberg et al.'s (2019) measures and results aligned with CONSORT standards for adequate data reporting (Moher et al., 2010; Schulz et al., 2010), expecting that there would be some lack of following CONSORT standards. In line with concerns raised about questionable measurement practices (Flake & Fried, in press), I examined the extent to which the authors provided evidence to support their dependent measures and examined relevant comments about measures provided in the NDA. I also examined the author COIs reported by Kingsberg et al. (2019) as well as any listed medical writing support.

Method

I examined data from the following three sources: a) bremelanotide's Food and Drug Administration NDA; United States Food and Drug Administration, 2019), b) clinicaltrials.gov protocol entries for the two Phase III bremelanotide trials (ClinicalTrials.gov, 2018a, 2018b), and c) the Kingsberg et al. journal article that reported data from both Phase III trials (Study 301 and Study 302) of bremelanotide (Kingsberg et al., 2019).

I also conducted a meta-analysis of efficacy and dropout data appearing in the FDA NDA and compared these outcomes to outcomes reported in Kingsberg et al. (2019). For continuous outcomes, data based on means and standard deviations were used to compute a standardized mean difference effect size. Effect sizes were weighted by their inverse variance when creating a pooled effect size (Hedges & Olkin, 1985). This was converted to Hedges' *d* to control for a small bias in the standardized mean difference effect size (Hedges & Olkin, 1985). In addition, where data reporting was sufficient, the raw difference in mean scores at posttest was analyzed, as this may provide useful information about benefits of treatment. A meta-analysis of two trials is certainly rather thin, but both trials were reasonably large and reported identical

methodology, thus rendering it sensible to pool them via meta-analysis.

For categorical outcomes, odds ratios, risk ratios, as well as number needed to treat (NNT) were calculated for efficacy outcomes and number needed to harm (NNH) was calculated for safety/tolerability outcomes. NNT represents the number of participants who would need to be treated with bremelanotide to gain one additional beneficial outcome which would not have been achieved had all patients taken placebo. NNH represents the number of participants who would need to be treated with bremelanotide to cause one additional harm which would not have occurred had all participants taken placebo. Comprehensive Meta-Analysis Version 2 software was used for analysis unless otherwise noted (Biostat, 2010). Heterogeneity was examined using the Q statistic. In addition, I^2 was used to report the amount of true heterogeneity relative to total effect size variability (Higgins et al., 2003). A random effects model was used for all analyses (DerSimonian & Laird, 1986). Although only two clinical trials with identical study designs were included, a random effects model was used because there is often variance across the many sites which comprise clinical trials (Kraemer & Robinson, 2005). For instance, different site investigators may recruit participants who vary in many ways and may interact with participants in different ways that could impact their scores on the dependent measures. NNT and NNH calculations were based on odds ratios rather than risk differences, as risk differences are subject to greater between-trial heterogeneity (Deeks, 2002). The baseline risk (needed for calculating NNT/NNH) was estimated by using the pooled event rate among placebo participants weighted by each study's sample size. NNT and NNH were calculated using Visual RX (Cates, n.d.).

Kingsberg et al. (2019) reported that after completing the phase III trials, participants were offered a chance to continue into an open-label phase of the trial. It seems logical that patients who both completed the acute phase and volunteered to continue into the open-label phase of the study perceived treatment to be both reasonably efficacious and tolerable. Thus, I used this as an overall measure of treatment preference.

The concordance of data reporting between Kingsberg et al. (2019) and CONSORT standards was examined. Kingsberg et al. (2019) stated that their paper followed Good Publication Practice (GPP3 – Battisti et al., 2015). GPP3 requires that clinical trials adhere to CONSORT data reporting standards. For continuous outcomes, CONSORT requires the following: a) summary statistics (means and standard deviations), b) report of the difference between group means and c) confidence interval for the difference between groups. For binary outcomes (e.g., treatment response), CONSORT requires a) the count of outcomes in each group, b) relative effect measures (e.g., either odds ratio or relative risk) with a confidence interval and c) absolute effect measure (risk difference) with a confidence interval.

Kingsberg et al. (2019) reported some data analyses as "integrated" across the two trials, meaning that data from the studies were pooled. Given that two separate studies were conducted, I treated data as coming from two separate studies in my meta-analytic calculations.

Results

Conflicts of Interest

Kingsberg et al. (2019) had four authors who worked for either the company that conducted the phase III trials (Palatin Technologies) or the company that was licensed to market bremelanotide in North America (AMAG Pharmaceuticals). AMAG has since divested its interest in bremelanotide, returning licensing rights to Palatin (AMAG Pharmaceuticals, 2020). The remaining four authors all have relevant financial conflicts of interest with AMAG and/or Palatin.

Changed Efficacy Outcomes

Several of the main problems discovered in my re-analysis of Kingsberg et al. (2019) are described briefly in Table 1. One main problem was the lack of reporting protocol-specified analyses. The clinicaltrials.gov study protocol for each trial indicated that 11 efficacy outcomes would be analyzed. Data from eight of these eleven outcomes (72.72%) were not reported in the Kingsberg et al. paper in a manner consistent with the clinicaltrials.gov protocol (Table 2). For seven outcomes, data were presented in terms of categorical outcomes by Kingsberg et al. but the clinicaltrials.gov protocol indicated that mean change would be analyzed. Kingsberg et al. (2019) provided no rationale for analyzing these as categorical measures. On two protocol-specified variables, FSFI total score and FSDS-DAO total score, in addition to categorical outcome analysis, data on the total scores (a continuous outcome) were vaguely described as positive by Kingsberg et al. (2019) without the provision of any data. As can be seen in Tables 3 and 4, Kingsberg et al. reported, in some form, results for 15 outcomes (one continuous and 14 categorical) which were not listed in the clinicaltrials.gov protocol entries.

One of the coprimary outcomes changed over time, with the FDA allowing the sponsor's request for satisfying sexual events (SSEs) to move from a coprimary to the key secondary outcome (United States Food and Drug Administration, 2019). This change occurred over a year after the trials had begun. Kingsberg et al. (2019) did not mention that this change occurred.

Efficacy Results: Coprimary Outcomes

My meta-analytic results (based on NDA data) on the two coprimary outcomes, the Female Sexual Function Index – Desire domain (FSFI-D; Rosen, 2000) and Female Sexual Distress Scale – Desire/Arousal/Orgasm #13 (FSDS-DAO #13; DeRogatis et al., 2008) can be seen in Table 3. Bremelanotide was superior to placebo by a small and statistically significant margin in terms of effect size. The advantage for bremelanotide on the FSDS-DAO #13 was 0.33 raw units. This question regarding frequency of being bothered by low sexual desire has five anchor points, each differing by one point on the scale: never (0), rarely (1), occasionally (2), frequently (3), and always (4). There is little literature about how to empirically interpret raw scores on the FSDS-DAO #13.

The FSFI-D is comprised of two items. One item inquires about frequency of feeling sexual desire/interest and the other

Table 1. Main areas of concern regarding Kingsberg et al. (2019).

Problem	Brief description	Why this is problematic
Most protocol-specified outcomes are unreported	8 of 11 protocol-specified efficacy outcomes are not reported in the manner specified in the protocol	-Decreased transparency -Unknown outcomes on most *a priori* outcomes leads to inadequate understanding of treatment efficacy -This violates CONSORT standards (Schulz et al., 2010)
Reporting of non-protocol specified efficacy outcomes	-15 efficacy outcomes not specified in the clinicaltrials.gov protocol are reported in Kingsberg et al. (2019)	-The post-hoc nature of these analyses limits confidence in their results -Some post-hoc analyses may have been a result of data dredging to find outcomes upon which bremelanotide demonstrated efficacy -Positive findings on *a priori* analyses are more convincing than positive findings on *post hoc* analyses
Several variables reported as showing favorable "trends" or as favoring treatment lack any numerical data	-Two continuous variables and four categorical variables were described as favorable without providing any quantification or statistical analyses	-Not providing summary statistics or statistical analyses renders these favorable outcomes highly questionable given their lack of transparency, which does not meet CONSORT standards (Schulz et al., 2010)
Dichotomizing outcomes from continuous outcomes without justification	-Post-hoc categorical outcomes were derived from cutoff scores on underlying continuous rating scales, including seven outcomes which were listed as continuous outcomes on the clinicaltrials.gov protocols.	-These dichotomous measures lack evidence of validity -A lack of validity evidence erodes confidence in the meaningfulness of these measures (Flake & Fried, in press; MacCallum et al., 2002) -Conversion of continuous outcomes to categorical runs risk of selecting cutoff points to maximize apparent efficacy (Altman & Royston, 2006; Kirsch & Moncrieff, 2007)
Lack of empirical justification for post-hoc measures	The authors provided no rationale for selection of any post-hoc measures	-Without a convincing rationale or evidence, the selected measures are of unclear validity (Flake & Fried, in press) -A lack of rationale for the post-hoc efficacy outcomes overlooks the potential lack of validity of these outcomes -CONSORT standards state that valid measures should be used (Moher et al., 2010)
Absolute benefit is incalculable for nearly all categorical analyses	Absolute benefit is reported for only one categorical outcome, whereas relative benefit was reported for all categorical outcomes	-Not reporting absolute benefits makes it impossible to tell how many people would need to receive treatment in order to derive additional treatment benefit. -CONSORT calls for reporting of both absolute and relative benefit (Schulz et al., 2010)
Number of dropouts due to adverse events is not reported by group	The total number of dropouts due to adverse events is provided, but this is not broken down by group.	-Readers of Kingsberg et al. are left unaware of the much higher dropout due to AE rate on bremelanotide versus placebo: Relative risk = 9.95, NNH = 6. -A lack of accurately reporting high dropout due to adverse events on bremelanotide provides false reassurance of the drug's tolerability. -CONSORT calls for clear reporting of dropouts and reasons for dropout in each group (Schulz et al., 2010)
Data reporting does not match CONSORT or GPP3 guidelines	CONSORT and GPP3 provide widely accepted standards for data reporting in clinical trials.	-Data reporting standards are intended to ensure adequate reporting of benefits and harms, while ensuring some level of transparency. Failure to follow these standards lowers confidence in the paper's conclusions. -The authors incorrectly stated that GPP3 was followed. GPP3 and *Obstetrics & Gynecology* author instructions state that CONSORT should be followed. Thus, the authors implicitly endorsed that CONSORT was followed, though it was not.
Change in coprimary measure is unreported	The number of sexually satisfying events (SSEs) was a coprimary measure, but was shifted to a secondary measure without disclosure	-All changes of outcomes should be reported to maximize transparency and reduce the chance of selecting primary measures based on their results. -Changing primary measures may or may not have been justified. Failing to disclose that a primary outcome was changed lacks transparency. Bremelanotide had no benefit on SSEs, which would seem more notable to readers if SSEs were a primary outcome.
Author and nonauthor contributions are unclear	Particularly in the face of other problems listed here, it is important that the roles of individual authors/contributors are reported for the sake of accountability. The name of the medical writer(s) hired by bremelanotide's sponsor is not listed in the paper.	-GPP3 states that author and nonauthor contributions should be clearly explained. Further, all authors and nonauthor contributors should be named (Battisti et al., 2015). - A lack of transparency makes it impossible to know who was responsible for the various problems listed elsewhere in this table. An unnamed medical writer from Phase Five Communications hired by bremelanotide's sponsor provided undefined "editorial support" for the paper. Phase Five's website makes claims such as "We sift through the client's raw data and polish it into the diamonds that make for great brands (Phase Five Communications, 2020)." In concert with the other concerns raised here, it is possible that commercial interests drove the way data were presented in a favorable manner for bremelanotide.

Table 2. Pre-specified efficacy outcomes listed in clinicaltrials.gov study protocols.

Outcome	Importance	How Presented in Kingsberg et al. (2019)	Kingsberg Analysis Matches Protocol?
FSFI-D: Questions 1 and 2	Primary	Effect size, approximate *p*-value for combined studies	Yes
FSDS-DAO: Question 13	Primary	Effect size, approximate *p*-value for combined studies	Yes
Number of SSEs	Secondary, changed from primary over one year after study began	Mean difference, exact *p*-value for combined studies	Yes
FSEP-R #3: Desire mean score	Secondary	Presented as categorical outcome	No
FSEP-R #4: Satisfaction with desire mean score	Secondary	Presented as categorical outcome	No
FSDS-DAO total score	Secondary	Presented as categorical outcome. Also, Kingsberg et al. state that "the FSDS total score highlights reduction in overall distress and parallels the overall improvement in the FSFI-D score." While "total score" implies a continuous measure, there was no quantification of FSDS-DAO results in terms of a continuous measure. In the discussion, the total FSDS-DAO total score is labeled a "supportive secondary endpoint" that, among others, provides "robust and consistent data (p. 906)" to support the efficacy of the drug.	No
FSFI total score	Secondary	Presented as categorical outcome. Also, Kingsberg et al. state that "total scores for FSFI" were used to assess "overall sexual function", but no continuous outcome data were provided on this measure. In the discussion, the total FSFI score is labeled a "supportive secondary efficacy endpoint" that, among others, provides "robust and consistent data" to support the drug's efficacy.	No
FSEP-R #6: Mean level of sexual arousal during SE	Secondary	Presented as categorical outcome	No
FSEP-R #7: Mean satisfaction with arousal during SE	Secondary	Presented as categorical outcome	No
FSDS-DAO #14: Mean time spent concerned by difficulty with sexual arousal	Secondary	Presented as categorical outcome	No
FSFI Arousal Domain #3- #6	Secondary	Presented as categorical outcome	No

inquires about the intensity of sexual desire/interest. Scores on each item range from 1 (very low) to 5 (almost always/always). The FSFI-D score is the combined score on the two items multiplied by 0.6. In my meta-analysis of NDA data, the difference favoring bremelanotide over placebo on the FSFI-D was .36 units, which when multiplied by the inverse of 0.6, generates a score of .602. This number represents the average raw score difference favoring bremelanotide when combining the two items on the FSFI-D. As with the FSDS-DAO #13, there is little evidence to guide how to interpret raw scores on the FSFI-D.

Efficacy Results: Secondary Outcomes

Tables 3 and 4 show results of the continuous and categorical outcomes, respectively. For 10 categorical outcomes, Kingsberg et al. (2019) reported some sort of quantitative analysis indicating superiority of bremelanotide over placebo. Four additional categorical outcomes were reported as showing a favorable "trend" for bremelanotide, with no numerical data provided. Similarly, two continuous outcomes were described as "supportive secondary endpoint[s]" that "provide robust and consistent data" in support of bremelanotide's efficacy without any numerical data. None of the favorable secondary efficacy outcomes resulted from data analyses matching the planned data analyses reported in the clinicaltrials.gov protocol. With one exception, the statistical analyses of positive categorical outcomes in Kingsberg et al. reported data solely in terms of relative difference between groups (odds ratios). On 9 of 10 statistical analyses of secondary categorical outcomes that favored bremelanotide, the numbers of participants who experienced beneficial outcomes in treatment and placebo groups were not reported; absolute treatment benefit was thus incalculable. On seven outcomes, the clinicaltrials.gov protocol described the *a priori* analysis in terms of mean change, but Kingsberg et al. reported these variables in terms of categorical outcomes. No rationale or validity data for these categorical outcomes were provided by the authors. As noted in Table 3, there were six secondary continuous outcomes mentioned by Kingsberg et al. (2019) upon which quantitative results were not provided (FSFI total, FSFI #16, FSDS-DAO total, FSDS-DAO #1, GAQ #3, EDQ #9). One such measure, General Assessment Questionnaire Item #3, was presented in a figure without providing exact numbers. Kingsberg et al. (2019) provided no citation or validity information for the GAQ. Further, the FDA NDA noted that the GAQ has not been validated (United States Food and Drug Administration, 2019).

Some rating scale items share the same number of ordinal rating points (e.g., they are scored on a 4-point rating scale). As shown in Table 4, Kingsberg et al. (2019) used different cutoff scores to define success for several individual items on the Female Sexual Encounter Profile-Revised scale, even though these items were each rated on a 4-point scale. The authors provide no description for why there should be different cutoff points for "improvement" on each of these items.

In terms of number of satisfying sexual events, bremelanotide provides no benefit (Table 3). Kingsberg et al. (2019) described a post-hoc analysis showing that a greater percentage of sexual events were satisfying on bremelanotide versus placebo (Table 4). However, the NDA mentions "At almost every

Table 3. Continuous efficacy outcomes.

Outcome (secondary unless listed as coprimary)	Study	Source	Prespecified outcome: Included in clinicaltrials.gov protocol entry?	d+ (95% CI for totals)	Raw units (95% CI for totals)	p(d+)	Q	I^2	p(Q)	Other Description and Notes
Current Analyses										
FSFI-D Desire (#1 and #2 combined): Coprimary outcome	301	FDA	Yes	.29	.30					
	302	FDA	Yes	.43	.42					
	Total	FDA	Yes	.35 (.21-.49)	.36 (.24-.48)	<.001	1.50	33.15%	.22	
FSDS-DAO #13: Coprimary outcome	301	FDA	Yes	.32	.37					
	302	FDA	Yes	.26	.29					
	Total		Yes	.30 (.18-.41)	.33 (.20 – .46)	<.001	0.26	0%	.61	
FSEP-R #10: Number of satisfactory sexual encounters	301	FDA	Yes	.05	.07					
	302	FDA	Yes	.04	.05					
	Total	FDA	Yes	.04 (−.07 – .16)	.06 (−.09 – .21)	.44	.01	0%	.93	
Elements of Desire Questionnaire (EDQ) Total	Total	FDA	No	*	*	*	*	*	*	Data are not valid[a]
Elements of Desire Questionnaire (EDQ) Items 1–9	Total	FDA	No	*	*	*	*	*	*	Data are not valid[a]
Analyses Reported in Kingsberg et al.										
FSFI-D Desire (#1 and #2 combined): Coprimary	301	Kingsberg	Yes	?	.30					
	302	Kingsberg	Yes	?	.42					
	Total	Kingsberg	Yes	.39 (?) for "integrated dataset" (?)	.35 (?)	?	?	?	?	
FSDS-DAO #13: Coprimary	301	Kingsberg	Yes	?	.37					
	302	Kingsberg	Yes	?	.29					
	Total	Kingsberg	Yes	.27 (?) for "integrated dataset"	.33 (?)	?	?	?	?	
FSFI #16 (Satisfaction with overall sexual life)	Total	Kingsberg	No	?	?	?	?	?	?	This item was used in a "sensitivity analysis," though the purpose of this analysis is unclear. No results are reported.
FSFI Total Mean Score	Total	Kingsberg	Yes	?	?	?	?	?	?	Data on this outcome were dichotomized. Also, data on the mean score were vaguely referenced as providing positive results.[b]
FSFI Arousal Domain (Items 3–6) mean change	Total	Kingsberg	Yes	?	?	?	?	?	?	Data on this outcome were dichotomized – see Table 4
FSDS-DAO #14 (time being concerned with sexual arousal difficulty)	Total	Kingsberg	Yes	?	?	?	?	?	?	Data on this outcome were dichotomized – see Table 4
FSDS-DAO Total Score	Total	Kingsberg	Yes	?	?	?	?	?	?	Data on this outcome were dichotomized. Also, data on the mean score were vaguely referenced as providing positive results.[c]
FSDS-DAO #1 (distress about sex life)	Total	Kingsberg	No	?	?	?	?	?	?	This item was used in a "sensitivity analysis," though the purpose of this analysis is unclear. No results are reported.

(*Continued*)

Table 3. (Continued).

Outcome (secondary unless listed as coprimary)	Study	Source	Prespecified outcome: Included in clinicaltrials.gov protocol entry?	d+ (95% CI for totals)	Raw units (95% CI for totals)	p(d+)	Q	I^2	p(Q)	Other Description and Notes
FSEP-R #10: Number of satisfactory sexual encounters	301	Kingsberg	Yes	.02	.1					I calculated effect size for study 301 calculated based on mean difference and *p*-value reported by Kingsberg et al.
	302	Kingsberg	Yes	?	0					Mean difference was zero in study 302, according to Kingsberg but *p*-value was .704. It was not stated which group had very slightly more events, so the effect size is incalculable.
	Total	Kingsberg	Yes	?	?	.				~~"integrated studies"~~
General Assessment Questionnaire #3 (perceived treatment benefit)	301	Kingsberg	No	?	?	?	?	?	?	Kingsberg et al. present the means graphically in their Figure 1. No exact means or standard deviations are presented.
	302	Kingsberg	No	?	?	?	?	?	?	Kingsberg et al. present the means graphically in their Figure 1. No exact means or standard deviations are presented.
	Total	Kingsberg	No	?	?	?	?	?	?	Kingsberg et al. present the means graphically in their Figure 1. No exact means or standard deviations are presented.
EDQ #9: Desire/Interest in sex	Total	Kingsberg	No	?	?	?	?	?	?	This item was used in a "sensitivity analysis," though the purpose of this analysis is unclear. No results are reported.

[a]The FDA noted the use of a daily recall version of the EDQ for seven consecutive days at several timepoints in the study. A 30-day recall version was also used. The FDA presents results when combining daily scores into a one-week average but notes that "there were large amounts of missing data for the daily diary version." Specifically, slightly more than a third of participants reported data for ≤ 3 of 7 days during the weeks when the EDQ daily version was to be scored by participants. Thus, I considered EDQ data to be invalid and did not analyze them. Further, the daily version was only administered intermittently throughout the study, so even participants who fully complied with completing the EDQ would have only reported data for four weeks of the 24-week study.

[b]Kingsberg et al. described this continuous measure being used to assess "overall sexual function" but no data comparing Bremelanotide (BRE) to Placebo (PLA) ~~was~~ provided. In the discussion, this measure is labeled a "supportive secondary efficacy endpoint" that, among others, provides "robust and consistent data" to support BRE's efficacy

[c]Kingsberg et al. mentioned that "overall … [low desire] associated distress" was assessed using the FSDS-DAO total score, but no data comparing mean levels of change on this measure are provided. In the discussion, they wrote that "the FSDS total score highlights reduction in overall distress and parallels the overall improvement in the FSFI-D score (906)." Further, they stated that the FSDS-DAO total score was a "supportive secondary endpoint" that, among others, provided "robust and consistent data (p. 906)" to support the efficacy of the drug.

Table 4. Categorical outcomes from bremelanotide Phase III trials.

Outcome: All secondary outcomes	Study	Data Source	Prespecified outcome: Included in clinicaltrials.gov entry?	Events/Total	OR (95% CI for totals), *p*-value	RR (95% CI for totals), *p*-value	NNT or NNH	Q	I^2 for OR	P(Q)	Other description/Notes[a]
Efficacy Outcomes											
Current Analyses											
FSFI-D Response: Improved by ≥ 1.2 and completed study	301	FDA	No	75/314 BRE 57/316 PLA							
	302	FDA	No	73/282 BRE 46/290 PLA							
	Total		No		1.61 (1.22–2.14), $p = .001$	1.46 (1.16–1.83), $p = .001$	13	0.82	0%	.36	
FSDS-DAO #13 Responder: Improved by ≥ 1 and completed study	301	FDA	No	116/314 BRE 98/316 PLA							
	302	FDA	No	93/282 BRE 93/290 PLA							
	Total		No		1.17 (0.92–1.49), $p = .19$	1.11 (0.95–1.31), $p = .19$	44	0.83	0%	.36	
Analyses Reported in Kingsberg et al											
FSFI Total Score Response (improvement by ≥ 4.2)	301	Kingsberg	No	?	2.12	?					
	302	Kingsberg	No	?	2.27	?					
	Total	Kingsberg	No	?	2.19 (1.70–2.83), $p = ?$	?	?	?	?	?	
FSFI response: "mean level of desire"[b]	301	Kingsberg	No	?	?	?					"Trend" toward greater response on BRE. Neither the cutoff score for a treatment response nor the items that constituted this measure were provided.
	302	Kingsberg	No	?	?	?					"Trend" toward greater response on BRE. Neither the cutoff score for a treatment response nor the items that constituted this measure were provided.
	Total		No	?	?	?	?	?	?	?	
FSFI response: "mean level of arousal"	301	Kingsberg	No	?	?	?					"Trend" toward greater response on BRE. Neither the cutoff score for a treatment response nor the items that constituted this measure were provided.
	302	Kingsberg	No	?	?	?					"Trend" toward greater response on BRE. Neither the cutoff score for a treatment response nor the items that constituted this measure were provided.
	Total		No	?	?	?	?	?	?	?	
FSFI response: mean satisfaction with level of desire	301	Kingsberg	No	?	?	?					"Trend" toward greater response on BRE. Neither the cutoff score for a treatment response nor the items that constituted this measure were provided.
	302	Kingsberg	No	?	?	?					"Trend" toward greater response on BRE. Neither the cutoff score for a treatment response nor the items that constituted this measure were provided.
	Total		No	?	?	?	?	?	?	?	

(*Continued*)

Table 4. (Continued).

Outcome: All secondary outcomes	Study	Data Source	Prespecified outcome: Included in clinicaltrials.gov entry?	Events/Total	OR (95% CI for totals), *p*-value	RR (95% CI for totals), *p*-value	NNT or NNH	Q	I^2 for OR	P(Q)	Other description/Notes[a]
FSFI response: mean satisfaction with level of arousal	301	Kingsberg	No	?	?	?					"Trend" toward greater response on BRE. Neither the cutoff score for a treatment response nor the items that constituted this measure were provided.
	302	Kingsberg	No	?	?	?					"Trend" toward greater response on BRE. Neither the cutoff score for a treatment response nor the items that constituted this measure were provided.
	Total		No	?	?	?	?	?	?	?	
FSFI Arousal Domain change of ≥ 0.6	301	Kingsberg	No	?	1.96	?					
	302	Kingsberg	No	?	2.0	?					
	Total		No	?	1.98 (1.56–2.50), *p* = ?	?	?	?	?	?	
FSDS-DAO #14 Time spent concerned by difficulty with sexual arousal improved by ≥ 1	301	Kingsberg	No	?	1.68	?					
	302	Kingsberg	No	?	1.83	?					
	Total	Kingsberg	No	?	1.75 (1.39–2.20), *p* = ?	?	?	?	?	?	
FSDS-DAO total score responder (improved by ≥ 10)	301	Kingsberg	No	?	2.37	?					
	302	Kingsberg	No	?	1.86	?					
	Total	Kingsberg	No	?	2.11 (1.66–2.68), *p* = ?	?	?	?	?	?	
FSEP-R #3: Level of sexual desire during SE improvement ≥ .25	301	Kingsberg	No	?	1.51	?					
	302	Kingsberg	No	?	1.08	?					
	Total	Kingsberg	No	?	1.28 (1.01–1.62), *p* = ?	?	?	?	?	?	
FSEP-R #4: Satisfaction with level of sexual desire associated with and during SE ≥ .50	301	Kingsberg	No	?	1.50	?					
	302	Kingsberg	No	?	1.46	?					
	Total	Kingsberg	No	?	1.48 (1.17–1.88), *p* = ?	?	?	?	?	?	
FSEP-R #6: Level of sexual arousal during SE ≥ .25	301	Kingsberg	No	?	2.15	?					
	302	Kingsberg	No	?	1.45	?					
	Total	Kingsberg	No	?	1.77 (1.40–2.25), *p* = ?	?	?	?	?	?	
FSEP-R #7: Satisfaction with arousal during SE ≥ .465	301	Kingsberg	No	?	1.78	?					
	302	Kingsberg	No	?	1.50	?					
	Total	Kingsberg	No	?	1.63 (1.28–2.07), *p* = ?	?	?	?	?	?	

(*Continued*)

Table 4. (Continued).

Outcome: All secondary outcomes	Study	Data Source	Prespecified outcome: Included in clinicaltrials.gov entry?	Events/Total	OR (95% CI for totals), *p*-value	RR (95% CI for totals), *p*-value	NNT or NNH	Q	I^2 for OR	P(Q)	Other description/Notes[a]
FSEP-R: Percentage of SEs rated as satisfactory	Total	Kingsberg	No	?	?	?	?	?	?	?	BRE: Improvement of 25.0% PLA: Improvement of 9.8% – $p < .001$. BRE: 64.6% of SEs rated as satisfactory at end of study PLA: 49.2% of SEs rated as satisfactory at end of study.
General Assessment Questionnaire #3 ≥ 5 (perceived treatment benefit)	301	Kingsberg	No	BRE: 183/314 PLA: 114/316	2.48	1.62					Numbers of events based on percentages provided in Kingsberg et al.: 58.3% vs. 36.1% in Study 301 and 58.2% vs. 35.4% in Study 302 for BRE and PLA.
	302	Kingsberg	No	BRE: 164/282 PLA: 103/290	2.52	1.64					
	Total				2.50 (1.98–3.15), $p< .001$	1.63 (1.43–1.85), $p <.001$	5	.007	0%	.94	
Safety/Tolerability Outcomes											
Current Analyses											
Discontinued for any reason	301	FDA		BRE: 134/324 PLA: 45/319	4.29	2.93					
	302	FDA		BRE: 130/303 PLA: 82/301	2.01	1.58					
	Total				2.92 (1.39–6.15), $p= .005$	2.13 (1.16–3.92), $p= .01$	5	8.41	88.10%	.004	
Discontinued due to adverse event	301	FDA		BRE: 60/324 PLA: 3/319							
	302	FDA		BRE: 55/303 PLA: 9/301							
	Total				11.98 (3.74–38.37), $p < .001$	9.95 (3.19–31.07), $p < .001$	6	2.93	65.83%	.09	
Chose to participate in open-label extension[c]	Total	Current analysis based on Kingsberg		BRE: 254/363 PLA: 430/493	0.34 (0.24–0.48), $p < .001$	0.80 (0.74–0.87) $p < .001$	6	N/A	N/A	N/A	Data reported only for combined studies, not for each study
Completed acute-phase study and agreed to open-label extension	Total	Current analysis based on FDA/ Kingsberg[d]		BRE: 254/627 PLA: 430/620	0.30 (0.24–0.38), $p < .001$	0.58(0.52–0.65), $p < .001$	4	N/A	N/A	N/A	
Reported by Kingsberg et al. (2019)											
Discontinued for any reason	301	Kingsberg		BRE: 137/327 PLA: 52/326							Reported only percentages. No statistical analysis was performed. BRE: 41.9% vs PLA: 16.0%.
	302	Kingsberg		BRE: 135/308 PLA: 87/306							Reported only percentages. No statistical analysis was performed. BRE: 43.8% vs PLA: 28.4%.
	Total				2.72 (1.42–5.19), $p = .002$	2.00 (1.19–3.37)	5	6.72	85.14%	.01	

[a]If no statistical description of a result was provided, the authors' description is provided here.
[b]It is unclear which specific item(s) comprised this measure.
[c]The denominator reflects only participants who completed the double-blind acute phase of the study.
[d]Using completer data from FDA and the data from Kingsberg et al. regarding participants who agreed to the open-label extension.
Note: BRE = Bremelanotide; PLA = Placebo.

visit, the [placebo] group had a higher number of [sexual] encounters (United States Food and Drug Administration, 2019, p. 144)". It is difficult to interpret this finding given that Kingsberg et al. did not report the number of reported sexual events and the number of satisfying events in each group.

Discontinuation Outcomes

In the abstract, Kingsberg et al. (2019) reported that "the safety profile was favorable" and that "Most treatment-related adverse events were related to tolerability and the majority were mild or moderate in intensity (p. 900)." Using data from the NDA, I found that rates of discontinuation were substantially higher for bremelanotide compared to placebo, with 42.1% of bremelanotide participants not completing a study compared to 20.48% of participants taking placebo (Table 4). The data on discontinuation rates differ slightly between the FDA NDA and the Kingsberg et al. article (Table 4). The present analysis used the number of participants in the study safety sample as the denominator (participants who were randomized and received at least one dose of drug or placebo). It appears that Kingsberg et al.'s calculations also included participants who were randomized but had not yet taken a dose of study drug or placebo during the randomized phase (their calculations match FDA calculations that included the randomized sample as the denominator). The present analysis operates under the assumption that it is more appropriate to only include participants who had taken a dose of treatment during the randomized phase, but in any case, the two analyses yield very similar results.

Kingsberg et al. (2019) listed dropouts due to adverse events in an appendix, and within the appendix, dropouts due to adverse events were reported only in the aggregate, not broken down by bremelanotide compared to placebo. This omission makes it impossible for a reader of the Kingsberg article to compare dropout rates due to adverse events between groups. It also does not follow CONSORT standards (Schulz et al., 2010). According to my meta-analysis based on the NDA, dropout rates due to adverse events were much higher for bremelanotide than placebo (Table 4), with a relative risk of 9.95 and an NNH of 6. There was some heterogeneity in this analysis, which is clearly explained by the placebo rate of dropouts due to AEs varying between 0.9% in Study 301 and 3.0% in Study 302. The rate of dropouts due to adverse events was highly consistent for bremelanotide: 18.52% and 18.15% in studies 301 and 302, respectively (Table 4).

Among participants who reached the end of the acute phase, more participants in the placebo group wanted to continue treatment in the open-label phase that followed the acute phase (87.22% vs. 69.97%). I defined treatment preference based on whether participants both completed the acute phase and agreed to continue into the open-label phase. On this measure, bremelanotide led to substantially lower persistence than bremelanotide: 69.35% for placebo versus 40.51% for bremelanotide (OR: 0.30, 95% CI = 0.24 – .0.38; NNH 4).

Adherence With CONSORT Standards

None of the nine continuous efficacy outcomes mentioned in Kingsberg et al. (2019) (three of which were used in sensitivity analyses and had no reported results) were reported according to CONSORT standards. However, the authors presented means and effect sizes for the two coprimary outcomes, the FSFI-D and FSDS-DAO #13, for the pooled dataset. They did not meet CONSORT standards because no standard deviations or confidence intervals were provided, although these outcomes were presented more transparently than other continuous outcomes in their paper. None of the 14 categorical measures with quantitative results were reported according to CONSORT standards. No categorical measure directly reported the number of responders and nonresponders. One categorical outcome (General Assessment Questionnaire Question #3 ≥ 5) reported percentages of responders in each group, from which I was able to calculate the number of responders, as well as an odds ratio and relative risk with appropriate confidence intervals. Dropout due to adverse events was not reported by group by Kingsberg et al. (2019), which is not in alignment with CONSORT standards.

CONSORT states that previously validated scales should be used as dependent measures whenever possible. Further, "Authors should indicate the provenance and properties of scales (Moher et al., 2010, p. 7)." The coprimary FSFI-D and FSDS-DAO #13 measures were the only two outcomes for which at least one citation of relevant psychometric qualities was provided. No other measure provided either a citation or any rationale for its reliability or validity; this falls short of CONSORT standards.

Efficacy Results: Excluded Outcomes

It was unknown exactly what was included as a secondary outcome in the FDA NDA, as the NDA stated that due to the key secondary outcome (satisfying sexual events) not showing a statistically significant advantage for bremelanotide, the other exploratory outcomes were generally not described further in the NDA, with the exception of data on the Elements of Desire Questionnaire (EDQ).

Two versions of the EDQ were used, one of which required the participant to recall relevant sexual desire/activity on a monthly basis. The other version was administered daily, but only during the week before the four clinical assessment points. Thus, even participants who completed each daily EDQ would provide data from only 4 weeks of the 24-week trial. Additionally, 31% of participants in Study 301 and 36% of patients in study 302 did not return EDQs with completed entries on four or more days of the weeks they were administered (United States Food and Drug Administration, 2019). Due to the high level of missing data on the EDQ and its infrequent administration during the trial, I did not consider it to be valid; it was thus not included in data analyses (for more explanation, see Unclear Meaning of Outcome Measures section). Data from the monthly version of the EDQ are not provided in either Kingsberg et al. (2019) or the NDA. Further, it does not appear that the EDQ was validated prior to the phase III bremelanotide trials; data from Phase III bremelanotide trials as presented in

conference abstracts are apparently the basis of the quite limited validity data that are currently available for this measure (Derogatis et al., 2020).

Defining Treatment Response

In the NDA, it was written that the "clinical meaningfulness" of treatment efficacy can be based, to an extent, on analyses of treatment response (p. 145). Bremelanotide's sponsor assembled an "Independent Anchor Assessment Committee" (IAAC) to operationally define treatment response. This committee determined that change scores of ≥ 0.6 on the FSFI-D and ≥ 1.0 on the FSDS-DAO Item 13 represented meaningful change. The FDA reviewer accepted the proposed 1.0 point change on the FSDS-DAO Item 13 as meaningful, but stated that improvement of ≥ 1.2 on the FSFI-D was a more sensible measure of meaningful change. Neither Kingsberg et al. (2019) nor the NDA describe the IAAC's workings in detail. However, a poster presentation funded by bremelanotide's sponsor sheds some light on the IAAC process (Revicki et al., 2018). A subset of 243 participants from studies 301 and 302 were asked: "did you benefit overall from the study medication and, if so, was this benefit enough to be meaningful to you?" (Revicki et al., 2018). Responses were categorized as follows: a) no benefit from study treatment, b) benefit from study treatment, but not a meaningful one, or c) meaningful benefit from study treatment. It was not reported how many of these 243 participants were taking bremelanotide as opposed to placebo. Among those who improved by ≥ 0.6 on the FSFI-D (the sponsor's definition of response, which was less stringent than FDA's definition), 23.2% said they had no benefit from treatment, and 12.1% reported a nonmeaningful benefit (Revicki et al., 2018). Improvement by ≥ 1.0 on FSDS-DAO #13 was the sponsor's and FDA's shared definition of response. Among those reaching this result, 31% reported no treatment benefit and 9.5% said they had a nonmeaningful benefit. These results show that, at best, response on the FSFI-D and FSDS-DAO #13 was poorly calibrated with treatment response as reported on the exit survey. This suggests that treatment response as defined by the sponsor may not align with treatment response as experienced by participants. In the Kingsberg et al. (2019) article, results for response on either the FSFI-D or FSDS-DAO #13 are not reported.

Instead, Kingsberg et al. claimed that " . . . the bremelanotide group showed significantly greater numbers of responders compared with placebo, thus demonstrating clinically meaningful benefits from bremelanotide treatment in alignment with FDA guidances (p. 904)." However, the authors do not state which "response" outcome(s) are being referenced. According to the FDA's definition of response on the FSDS-DAO #13, bremelanotide did not outperform placebo. For those who met the sponsor's definition of response on the FSFI-D (a less stringent definition than that adopted by the FDA), 35% said on an exit survey that they had either no treatment benefit or a non-meaningful benefit. Using the FDA's FSFI-D response definition (improving by ≥ 1.2 points and completing the trial), treatment benefit was very small, with an NNT of 13 (Table 4).

Unclear Meaning of Outcome Measures

On the coprimary outcome of FSFI-D change, the current analysis calculated an effect size of .35, whereas Kingsberg et al. reported an effect size of .39. While this seems to indicate some degree of treatment efficacy, it is also important to consider what the FSFI-D actually represents. Factor analytic studies of the FSFI have mainly not found that desire is an independent domain (Neijenhuijs et al., 2019). Rather, such studies have typically found that the two FSFI desire items best fit alongside the four FSFI arousal items into a shared domain of desire and subjective arousal. In the FSFI's initial validation study, Rosen (2000)'s factor analysis did not support the creation of a "desire" domain. Rather, the FSFI-D domain was included due to "clinical consideration" (Rosen, 2000, p. 198), as a "panel of experts" concluded that splitting these domains "would provide greater ability to assess treatment specificity" (Rosen, 2000, p. 203). A review of the FSFI's properties suggested that the arousal and desire domains should be merged based on findings from various studies which have examined the FSFI's structure (Neijenhuijs et al., 2019). If desire is not actually a separate domain, then the FSFI-D should not be used to "assess treatment specificity", since the FSFI-D itself lacks specificity.

The FDA NDA states that "the FSFI desire domain (and with a 28-day recall) was not an optimal measure of desire (United States Food and Drug Administration, 2019, p. 339)." The FDA NDA noted that measuring treatment efficacy over a 28-day recall does not logically map onto a treatment taken acutely to purportedly boost one's sexual desire prior to a singular sexual encounter. Further, "the FDA considers the evidence to support the content validity of the FSFI to be limited (United States Food and Drug Administration, 2019, p. 118)." Authors of a recent systematic review of the FSFI also expressed concerns about the instrument's content validity (Neijenhuijs et al., 2019). It is also worth reiterating that the FSFI-D mapped poorly onto the exit survey interview question assessing meaningful change, with 35% of "responders" (according to the sponsor's definition) indicating that they had either no treatment benefit or a non-meaningful benefit.

The FDA allowed Palatin to conduct the Phase III trials using the FSFI-D but also requested to examine data from the EDQ to bolster the FSFI. As noted earlier, both a daily and monthly version of the EDQ were used. The daily version was infrequently used in the study and even less frequently completed, making it an unreliable outcome. Data on the monthly version are unavailable. This seems particularly problematic given that the FDA stated that the daily version of the EDQ "may bridge and give confidence for the monthly EDQ and subsequently the 28-day recall of the FSFI" (United States Food and Drug Administration, 2019, p. 119). Even if the EDQ would have been regularly completed, there is very little research to substantiate the validity of the EDQ (Clayton et al., 2018; Derogatis et al., 2020).

The General Assessment Questionnaire (GAQ) item 3 was the only secondary categorical outcome for which I was able to calculate the number of responders and non-responders in each group. It generated an NNT of 5 in favor of bremelanotide. However, Kingsberg et al. (2019) provided no citation for the

GAQ. The FDA NDA describes the GAQ as an outcome that "[has] not been validated" (United States Food and Drug Administration, 2019, p. 38). One study performed preliminary statistical validation of the GAQ as an outcome measure in HSDD based on results from a Phase II study of bremelanotide (Althof et al., 2019). This validation was based only on examining the relation of the GAQ to items, subscales, and total scores on the FSFI-D and FSDS-DAO. Such analysis is incapable of determining whether the GAQ can provide additional information beyond what can be obtained from these other instruments already included in Kingsberg et al. (2019). Further, this validation is quite preliminary. The GAQ was also not listed as a measure on the clinicaltrials.gov study protocol.

The FSDS-DAO #13 includes only a single rating scale item regarding how much a woman is bothered by her low sexual desire. It does not seem reasonable to expect that any one-item measure of distress would be particularly comprehensive or reliable. One study found that 14 of 25 women with HSDD indicated that item 13 covered all of their concerns related to low sexual desire (DeRogatis et al., 2011). The small sample size is concerning. Also, the fact that nearly half of the women found it did not cover all of their desire-related concerns suggests the measure is not comprehensive. Another study found that the test-retest reliability of item #13 was substantially lower than the reliability of the full scale FSDS, which again is what one would expect from a one-item measure (DeRogatis et al., 2008). Further, as noted previously, over 40% of those who "responded" on item 13 according to the sponsor indicated in an exit interview that they either did not have a response or that they had a nonmeaningful treatment response.

On some items of the Female Sexual Encounter Profile-Revised (FSEP-R), bremelanotide appeared to generate positive outcomes. Kingsberg et al. (2019) cited no evidence of this measure's validity. As with the other categorical measures reported by Kingsberg et al. (2019), it seems these outcomes were concocted post-hoc. In addition, no citation for the reliability or validity of this measure was provided by Kingsberg et al. (2019). An earlier trial of bremelanotide also used the FSEP-R. In reporting the outcomes of the trial, Clayton et al. (2016) provided one reference for the FSEP-R, a paper by Ferguson (2002), who briefly mentioned a few outcome measures, including the FSEP. He stated "The utility of all of these instruments has yet to be demonstrated in [female sexual dysfunction] (Ferguson, 2002, p. 82)." This does not reassure readers of the validity of the FSEP-R.

Editorial Support and Author Roles

Author instructions for *Obstetrics & Gynecology*, the journal in which the Kingsberg et al. paper was published, state "All persons who contributed to the work reported in the manuscript, but not sufficiently to be authors, must be acknowledged in a separate paragraph on the title page of the manuscript (Obstetrics & Gynecology, 2020)." In Kingsberg et al., there is a brief acknowledgment that Phase Five Communications provided "editorial support in the preparation of this manuscript," paid for by AMAG Pharmaceuticals, which was licensed to market bremelanotide in North America at the time of the manuscript's publication (Kingsberg et al., 2019, p. 899). No specific author from Phase Five is named. Not naming the writer(s) is in violation of journal standards.

The authors state that their paper followed GPP3, which states that author contributions, as well as contributions from nonauthors should be clearly described in the manuscript. GPP3 adds that all authors should also have access to relevant study data and the study protocol (Battisti et al., 2015). The authors thus should have been aware that they were not reporting data in accordance with protocol-specified statistical analyses. On a related note, GPP3 states that the sponsor should provide "all prespecified primary and secondary outcomes" to authors. Further, GPP3 states that "relevant contributions from persons who did not qualify as authors should also be disclosed (Battisti et al., 2015, p. 463)."

Obstetrics & Gynecology adheres to CONSORT standards. The Kingsberg et al. (2019) article was accepted after revisions made following one round of peer review. To promote transparency, *Obstetrics & Gynecology* provides peer review comments online. The paper was reviewed by three peer reviewers, a statistical reviewer, an associate editor, and the editorial office (Obstetrics & Gynecology, 2019). In their comments, no reviewer described comparing the submitted paper to the underlying clinicaltrials.gov entry. Further, reviewer comments about transparent data reporting were minimal. One reviewer called for reporting some quantification of "the magnitude of difference in sexual desire and sexual distress." Another reviewer called for providing the number of satisfying sexual events rather than just listing the analysis as not statistically significant. A reviewer called for providing confidence intervals in a figure. In the peer review, nobody requested that the authors report all outcomes in an appropriate manner that aligned with CONSORT standards (Obstetrics & Gynecology, 2019).

I submitted a version of this paper to *Obstetrics & Gynecology*. It contained the same data analyses and reached the same conclusions. The wording and organization differed somewhat based on the lower word count allowed by *Obstetrics & Gynecology*. One day after submission, the paper was rejected by *Obstetrics & Gynecology* after review by the editor and an editorial board member, with the following rationale: "Unfortunately, we can only publish a fraction of the papers received. Many submissions represent sound work, but space permits us to publish only those ranked highest." No specific comments about my paper were provided.

Discussion

Questionable Research and Measurement Practices

On the coprimary outcome measures (mean change on FSFI-D and FSDS-DAO #13), bremelanotide offers modest benefits over placebo. According to Kingsberg et al. (2019), several post-hoc categorical measures of treatment response demonstrated treatment benefits. However, a) these measures were not in accordance with the clinicaltrials.gov protocols' statistical analysis plans and b) no empirical justification was provided for the cutoff points used to determine "treatment response" on these various outcomes, and c) most protocol-specified outcomes were

not reported by Kingsberg et al. (2019). These are examples of questionable research and measurement practices (Flake & Fried, in press; John et al., 2012). It is concerning that the secondary efficacy outcomes were apparently derived post-hoc; this may be an example of "torturing the data" to extract the most positive spin on efficacy (Mills, 1993). Further, the continuous data captured on most mental health rating scales does not transform logically into dichotomous categories. If such conversions are made, they should be done in conjunction with cited and clearly described supportive evidence (Altman & Royston, 2006; Flake & Fried, in press; Kirsch & Moncrieff, 2007; MacCallum et al., 2002). There are several reasons to be skeptical of bremelanotide's purported benefits on these secondary efficacy outcomes.

According to FDA's definition of treatment response, bremelanotide offered either a very modest benefit (FSFI-D) or no benefit (FSDS-DAO #13). Kingsberg et al. (2019) did not report these findings. Of concern, Kingsberg et al. also failed to report the number of participants who dropped out due to adverse events by group, making it impossible for readers to ascertain ~~the much higher discontinuation rate on bremelanotide~~. The benefits described in the measures reported in the Kingsberg et al. article are likely greater than the benefits on the protocol-listed outcomes, in keeping with the wider literature on publication bias and selective outcome reporting in both drug industry trials (e.g., Jureidini et al., 2016; Le Noury et al., 2015; Roest et al., 2015; Ross et al., 2009; Spielmans & Parry, 2010; Turner, 2013; Turner et al., 2008) and "irreproducible science" more generally (Border et al., 2019; Bradley et al., 2017; Open Science Collaboration, 2015; Simmons et al., 2011). Clinicians, patients, and researchers should not read the main journal article describing clinical trial results and remain unaware of the results on the protocol-listed outcomes; CONSORT standards clearly call for reporting data on all prespecified outcomes (Schulz et al., 2010).

It is concerning that the peer review process failed to catch many ways in which Kingsberg et al. (2019) did not meet CONSORT standards (Obstetrics & Gynecology, 2019). I am not claiming that peer review served no purpose or resulted in no improvements to the initial paper. The published version may indeed represent a much-improved manuscript. Even if this is the case, the review process did not catch easily noticeable violations of the CONSORT standards to which *Obstetrics & Gynecology* adheres. GPP3 calls for the names of medical writers to be disclosed. GPP3 backs the use of a contributorship method of describing who did what; in the case of the Kingsberg et al. (2019) paper, this may resolve some ambiguity over who bears responsibility for some of the aforementioned problems in data reporting.

Researchers who do not clearly describe their measures, why they were selected, and provide evidence of their validity display a "measurement shmeasurement" approach to selecting dependent variables (Flake & Fried, in press). Such problems are widespread. Use of questionable, nontransparent measurement practices by Kingsberg et al. (2019) decreases faith in the authors' conclusions that bremelanotide demonstrated clear treatment benefit. Indeed, one might argue that their results provide more questions than answers. Here is just one of many potential examples: On the outcome of FSEP-R item 7 (satisfaction with sexual arousal) improving by greater than 0.465 points, bremelanotide outperformed placebo to a statistically significant extent (OR = 1.63, 95% CI: 1.28–2.07). On what empirical basis was this cutoff of 0.465 points selected? How many patients would need to be treated with bremelanotide to achieve one additional benefit? What evidence of validity exists for various cutoff points on this item? Why was this rating scale item transformed to a dichotomous measure? Why was this item analyzed separately from the total rating scale score?

Though quite commonly used in industry-supported journal articles, the mention of "editorial assistance" or "editorial support" provides no clarity as to what the medical writer(s) did in preparing the paper. Industry-supported clinical trials are typically designed by drug firms, who then analyze their own data (Matheson, 2016b; Sismondo, 2007, 2018; Sismondo & Nicholson, 2009). In developing journal articles which report clinical trial results, the involvement level of academic "authors" ranges from nominal to substantial. In many cases, the first draft of such manuscripts is drafted by a medical writer hired by the drug's sponsor (Healy & Cattell, 2003; Matheson, 2016b; McHenry & Amsterdam, 2019; McHenry & Jureidini, 2008; Ross et al., 2008). For instance, internal documents from the antidepressant paroxetine's manufacturer detail how a medical writer was in fact the key author of two manuscripts which mainly featured post-hoc analyses to paint an overly positive picture of drug efficacy while also minimizing the reporting of risks (Jureidini et al., 2008; McHenry & Amsterdam, 2019). Some people claim that a footnote acknowledging "editorial support" is sufficient to nullify any charges of ghostwriting. The Merriam-Webster dictionary states that ghostwriting is "to write for another who is the presumed or credited author" (Merriam-Webster, 2020). Suppose that a medical writer wrote a substantial portion – perhaps including the first draft – of a manuscript. Further supposing that the very substantial writing by the medical writer is not clearly described, this would tightly align with the dictionary definition of ghostwriting. Alastair Matheson, former medical writer, has aptly noted that "The 'problem' with ghostwriting is not secrecy but inadequate communication to readers about how the text was developed" (Matheson, 2016a, p. 1).

In the absence of any definition of "editorial support" in the Kingsberg et al. article, material from Phase Five's website appears relevant. Phase Five's main webpage states "We sift through the client's raw data and polish it into the diamonds that make for great brands." (Phase Five Communications, 2020). In a promotional piece that accompanies an article coauthored by two members of Phase Five Communications, it is stated that "Wendy Balter's [long-time Phase Five President] team of powerhouse conceptual alchemists transforms scientific base metal into strategic pure gold via exceptional marketing initiatives, medical meetings, and manuscripts. Connected with the industry's top opinion leaders and marketers, Phase Five's experienced PhDs and MDs understand how to energize your data with precious meaning. The result: powerful marketing programs to drive your brand to unexpected heights" (Phase Communications, n.d.). In addition, Phase Five states that "Our teams enjoy shaking up how to

look at product data … (Phase Five Communications, 2020)." Such clear discussion of Phase Five's business might be more informative about the role of its writers than a vague "editorial support" acknowledgment in the Kingsberg et al. article.

HSDD and Its Corporate Management

The current analysis mainly focuses on the unimpressive results of the two phase III trials of bremelanotide along with problematic data reporting in the journal article by Kingsberg et al. (2019). However, focusing solely on problems with the clinical trials runs the risk of unintentionally reifying the validity of HSDD (Hyman, 2010; Jutel, 2010). Indeed, the DSM-5 creation of female sexual interest/arousal disorder was an attempt to make the diagnostic manual better reflect the underlying, evolving science of female sexual functioning (Brotto, 2010; Graham et al., 2014). The lack of specifying symptom duration, questionable validity for the lack of sexual fantasies as a diagnostic criterion, difficulty in disentangling individual sexual problems from relational problems, and the failure to consider cultural influence (including the pressure on women to satisfy the sexual desires of their male partners) in the experience of sexuality all render HSDD as a problematic entity.

The role of the pharmaceutical industry in promoting HSDD has been cogently documented (Graham et al., 2017; Jutel, 2010; Moynihan, 2003; Tiefer, 2006). In order to market the idea of widespread female sexual dysfunction, epidemiological studies have been misinterpreted as showing that over 40% of women suffer from sexual dysfunction, with low desire often cited as occurring in at least 10% of women (Meixel et al., 2015). Laumann et al.'s (1999) study of sexual dysfunction prevalence in the United States has been cited over 6400 times (according to Google Scholar). The study found 43% of women experienced at least one symptom of "sexual dysfunction", but did not assess whether experiencing symptoms (including a lack of desire for sex) was associated with distress. Prevalence rates of sexual disorders decrease substantially as more stringent definitions of disorder are implemented. For instance, the National Survey of Sexual Attitudes and Lifestyles in the United Kingdom (NATSAL-3) found that 6.5% of a nationally representative sample of sexually active women experienced a lack of sexual interest and arousal, which the authors used as a rough proxy measure for symptom criteria for DSM-5's FSIAD. In the next step, the authors found that only 9.1% of women who reported these symptoms (0.6% of the total sample) met all of the following criteria: a) six-month minimum symptom duration, b) occurrence of symptoms "very often" or "always", and c) and being "fairly" or "very" distressed by symptoms (Mitchell et al., 2016). Their measure did not map exactly onto DSM-5 FSIAD criteria and they could not rule out other medical problems or relational problems as causing sexual problems. But the main point – that requiring substantial distress, symptom duration, and symptom frequency leads to much lower prevalence estimates – is well worth considering.

"Condition branding" refers to conveying the importance of a medical entity for marketing purposes, emphasizing the seriousness of a condition and the "unmet need" for treatment which purportedly benefits those who suffer from it (and also benefits those who sell treatments) (Angelmar et al., 2007). HSDD has been promoted through materials funded by the sponsors of pharmaceutical treatments for the condition. For example, sponsored continuing medical education materials (CME) have claimed that HSDD is underdiagnosed and undertreated, and can be diagnosed quickly using rating scales and/or screening measures – even among healthcare providers who lack specialty training in sexuality (Meixel et al., 2015). Treatment for HSDD is often recommended in such CME. Sprout, the sponsor of flibanserin, hired a consultant that created the "Even the Score" campaign, which pointed to a lack of treatments for female sexual dysfunction (Graham et al., 2017; Segal, 2015, 2018; Tavernise & Pollack, 2015). The campaign claimed that men had access to 26 FDA-approved treatments for male sexual dysfunction, yet no similar products were available for women. This might be considered misleading in that many of these 26 products were various formulations of testosterone, and there is no FDA-approved treatment for low sexual desire in males (Gellad et al., 2015). Nonetheless, the Even the Score website pointed to this inequity and noted that "there is still a long way to go before we achieve true gender equity in sexual health – and Even the Score will be there every step of the way" (Hogenmiller et al., 2017). Yet once flibanserin was FDA-approved for treating HSDD, Even the Score stopped producing content and eventually disappeared, with the score apparently evened by the drug's approval and whatever revenue could be generated from its sales.

The poorly defined symptoms of HSDD lend themselves to condition branding. Common and somewhat vaguely defined symptoms have helped to increase "awareness" and rates of diagnosis for conditions such as depression (Cosgrove et al., 2020), bipolar spectrum disorder (Healy, 2006; Paris, 2009; Spielmans, 2009), and social anxiety disorder (Lane, 2008). In some instances, of course, these diagnoses have led to people receiving treatment that has offered substantial benefit. But "awareness" of vaguely defined conditions can also lead to overdiagnosis and overtreatment and medicalize normal human experiences (Frances, 2014; Horwitz & Wakefield, 2007; Paris, 2015; Schwarz, 2016).

The corporate appropriation of feminist language to encourage diagnosis and treatment of HSDD is an interesting tactic. Even the Score and some advocates of treating HSDD with medication have portrayed seeking diagnosis and treatment for HSDD as empowering for women, who now have viable medications to treat their heretofore overlooked yet highly disabling medical condition (Goldstein, 2009; Graham et al., 2017; Tavernise & Pollack, 2015). Such language might be justified if women were being given access to a treatment that generally demonstrated clear benefit. Yet if women are to make a rational choice regarding treatment, they should be aware of the small degree of bremelanotide's efficacy, that the protocol-specified outcomes of bremelanotide are mostly unknown, and that participants would rather take a placebo than bremelanotide. Corporate-friendly feminist narratives are notably short on such details.

Limitations

The present analysis is limited in several ways. First, only two Phase III trials of bremelanotide were analyzed. Perhaps additional trials of bremelanotide would yield differing results. There is at least one other placebo-controlled trial of bremelanotide from an earlier phase in its development (Clayton et al., 2016). However, as a) FDA considers phase III trials to be "pivotal" in determining whether to approve a drug, b) usable data from the FDA NDA from the phase III (but not earlier phase) trials are available, and c) the phase II trial included women with DSM-IV diagnoses of female sexual arousal disorder, whereas Kingsberg et al. (2019) excluded participants with any female sexual dysfunction other than HSDD, only the two phase III trials were included in the current re-analysis. Nonetheless, the Phase II trial is briefly described for the sake of completeness (Clayton et al., 2016). The study used three different dosages, with one group receiving the 1.75 mg dose later used in the Phase III trials. Seven protocol-specified outcomes were listed in the study's clinicaltrials.gov entry (Palatin Technologies, 2014), three of which were not reported in the Clayton et al. (2016) article. A subset of outcomes were reported among participants with either an exclusive or primary HSDD diagnosis. Briefly, for the 1.75 mg dose, Clayton et al. found statistically significant efficacy for bremelanotide on three of five reported outcomes among patients with either an exclusive or primary HSDD diagnosis.

Due to various demographic characteristics as well as study inclusion and exclusion criteria, participants in the current meta-analysis may not be representative of patients seen in some clinical practice settings. Participants in the phase III bremelanotide clinical trials were American or Canadian, 85% of whom were Caucasian, with an average age of 38 years old. The generalizability of the evidence regarding bremelanotide's efficacy and tolerability is largely unknown. While meta-analysis offers a standardized method of data analysis, results may be interpreted in various ways. The present findings strongly suggest that bremelanotide's Phase III trial results paint a picture of very limited treatment efficacy and demonstrate that patients clearly prefer placebo over bremelanotide. However, other interpretations of efficacy and tolerability data are welcome, particularly if they are based on sound empirical and logical foundations.

Conclusion

Bremelanotide appears to offer modest benefits on the FSFI-D and FSDS-DAO #13. However, patients preferred taking placebo over bremelanotide, in terms of both a) much lower dropout rates and b) a higher likelihood of desiring to participate in the open-label extension phase. The frequent mismatch between outcomes reported in Kingsberg et al. and outcomes reported in the clinicaltrials.gov study protocols raises questions about the transparency of data reporting. Describing the treatment benefits of bremelanotide is challenging given that: a) outcomes on most protocol-specified outcome measures ~~is~~ are unknown; b) most reported efficacy outcomes were apparently derived post-hoc; c) most definitions of "responders" were derived from cutoff points lacking supporting evidence; and d) the numbers of participants who experienced "response" on nearly all categorical measures in Kingsberg et al. (2019) is unknown, making it impossible to calculate absolute treatment benefit. Both outcome selection and outcome reporting in Kingsberg lacked adherence to widely accepted CONSORT standards.

More succinctly, bremelanotide's benefits on mainly incompletely reported post-hoc measures of questionable validity fail to impress. Full reporting of data from all *a priori* measures and a convincing explanation of the empirical rationale behind the post-hoc measures would provide a clearer picture of bremelanotide's efficacy. In the interests of transparency, a clearer description of the authors' contributions, including the work of the anonymous contracted writer(s) who provided "editorial support" is also needed. Based on currently available evidence from the Phase III bremelanotide trials, it appears that patients prefer placebo over bremelanotide and that bremelanotide offers little benefit for women diagnosed with HSDD. My conclusions differ substantially from those reached in the article supported by bremelanotide's sponsor (Kingsberg et al., 2019), in which questionable research and measurement practices obfuscated the reporting of bremelantoide's efficacy and tolerability.

Acknowledgments

I have holdings in Vanguard Healthcare, a mutual fund that invests heavily in pharmaceutical firms. I thank Jaya Stenquist for her thoughtful discussion and background research regarding the medical and psychiatric conceptualizations of female sexuality. I thank Lori Brotto, Cynthia Graham, John Sakaluk, and an anonymous reviewer for their helpful comments on prior versions of this manuscript.

References

Althof, S., Derogatis, L. R., Greenberg, S., Clayton, A. H., Jordan, R., Lucas, J., & Spana, C. (2019). Responder analyses from a phase 2b dose-ranging study of bremelanotide. *The Journal of Sexual Medicine*, *16*(8), 1226–1235. https://doi.org/10.1016/j.jsxm.2019.05.012

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ*, *332*(7549), 1080. https://doi.org/10.1136/bmj.332.7549.1080

AMAG Pharmaceuticals. (2020, July 27). *AMAG pharmaceuticals completes divestment of women's health assets – AMAG pharmaceuticals*. AMAG Pharmaceuticals. https://www.amagpharma.com/news/amag-pharmaceuticals-completes-divestment-of-womens-health-assets/

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Angelmar, R., Angelmar, S., & Kane, S. (2007). Building strong condition brands. *Journal of Medical Marketing*, *7*(4), 341–351. https://doi.org/10.1057/palgrave.jmm.5050101

Armstrong, D. (2006). Medical reviews face criticism over lapses. *Wall Street Journal*, p. B1.

Battisti, W. P., Wager, E., Baltzer, L., Bridges, D., Cairns, A., Carswell, C. I., Citrome, L., Gurr, J. A., Mooney, L. A., Moore, B. J., Peña, T., Sanes-Miller, C. H., Veitch, K., Woolley, K. L., & Yarker, Y. E. (2015). Good publication practice for communicating company-sponsored medical research: GPP3. *Annals of Internal Medicine*, *163*(6), 461–464. https://doi.org/10.7326/M15-0288

Biostat. (2010). *Comprehensive meta-analysis*. Biostat.

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No support for historical candidate gene or candidate gene-by-interaction hypotheses for

major depression across multiple large samples. *American Journal of Psychiatry*, *176*(5), 376–387. https://doi.org/10.1176/appi.ajp.2018.18070881

Bradley, H. A., Rucklidge, J. J., & Mulder, R. T. (2017). A systematic review of trial registration and selective outcome reporting in psychotherapy randomized controlled trials. *Acta Psychiatrica Scandinavica*, *135*(1), 65–77. https://doi.org/10.1111/acps.12647

Brotto, L. A. (2010). The DSM diagnostic criteria for hypoactive sexual desire disorder in women. *Archives of Sexual Behavior*, *39*(2), 221–239. https://doi.org/10.1007/s10508-009-9543-1

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436. https://doi.org/10.1126/science.aaf0918

Cates, C. J. (n.d.). *Visual RX, Version 3*. Cates, C. nntonline.net/visualrx

Chan, A. W., Hrobjartsson, A., Haahr, M. T., Gotzsche, P. C., & Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *The Journal of the American Medical Association*, *291*(20), 2457–2465. https://doi.org/10.1001/jama.291.20.2457

Clayton, A. H., Althof, S. E., Kingsberg, S., DeRogatis, L. R., Kroll, R., Goldstein, I., Kaminetsky, J., Spana, C., Lucas, J., Jordan, R., & Portman, D. J. (2016). Bremelanotide for female sexual dysfunctions in premenopausal women: A randomized, placebo-controlled dose-finding trial. *Women's Health*, *12*(3), 325–337. https://doi.org/10.2217/whe-2016-0018

Clayton, A. H., Segraves, R. T., & Pyke, R. E. (2018). A brief version of the Sexual Interest and Desire Inventory–Female with 4 core items. *The Journal of Sexual Medicine*, *15*(9), 1370–1371. https://doi.org/10.1016/j.jsxm.2018.07.007

ClinicalTrials.gov. (2018a, October 2). *A phase 3, randomized, double-blind, placebo-controlled, parallel-group trial with an open-label extension*. National Library of Medicine. https://clinicaltrials.gov/ct2/show/NCT02333071

ClinicalTrials.gov. (2018b, October 2). *A phase 3, randomized, double-blind, placebo-controlled, parallel-group trial with an open-label extension*. National Library of Medicine. https://clinicaltrials.gov/ct2/show/NCT02338960

CONSORT. (2020). *Endorsers: CONSORT transparent reporting of trials*. Consolidated Standards of Reporting Trials Group. http://www.consort-statement.org/about-consort/endorsers

Cosgrove, L., Morrill, Z., Yusif, M., Vaswani, A., Cathcart, S., Troeger, R., & Karter, J. M. (2020). Drivers of and solutions for the overuse of antidepressant medication in pediatric populations. *Frontiers in Psychiatry*, *11*(11), 17. https://doi.org/10.3389/fpsyt.2020.00017

Deeks, J. J. (2002). Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine*, *21*(11), 1575–1600. https://doi.org/10.1002/sim.1188

DeRogatis, L., Clayton, A., Lewis-D'Agostino, D., Wunderlich, G., & Fu, Y. (2008). Validation of the Female Sexual Distress Scale-Revised for assessing distress in women with hypoactive sexual desire disorder. *The Journal of Sexual Medicine*, *5*(2), 357–364. https://doi.org/10.1111/j.1743-6109.2007.00672.x

DeRogatis, L., Pyke, R., McCormack, J., Hunter, A., & Harding, G. (2011). Does the Female Sexual Distress Scale-Revised cover the feelings of women with HSDD? *The Journal of Sexual Medicine*, *8*(10), 2810–2815. https://doi.org/10.1111/j.1743-6109.2011.02385.x

Derogatis, L. R., Revicki, D. A., & Clayton, A. H. (2020). Instruments for screening, diagnosis, and management of patients with generalized acquired hypoactive sexual desire disorder. *Journal of Women's Health*, *29*(6), 806–814. https://doi.org/10.1089/jwh.2019.7917

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry*, *168*(10), 1041–1049. https://doi.org/10.1176/appi.ajp.2011.11020191

Fava, G. A. (2016). The hidden costs of financial conflicts of interest in medicine. *Psychotherapy and Psychosomatics*, *85*(2), 65–70. https://doi.org/10.1159/000442694

Ferguson, D. M. (2002). Clinical trial development in female sexual dysfunction. *Journal of Sex & Marital Therapy*, *28*(sup1), 77–83. https://doi.org/10.1080/00926230252851212

Flake, J. K., & Fried, E. I. (in press). Measurement shmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*. https://doi.org/10.1177/2515245920952393

Frances, A. (2014). *Saving normal: An insider's revolt against out-of-control psychiatric diagnosis, DSM-5, big pharma, and the medicalization of ordinary life*. William Morrow Paperbacks.

Fugh-Berman, A. J. (2010). The haunting of medical journals: How Ghostwriting Sold "HRT." *PLoS Medicine*, *7*(9), e1000335. https://doi.org/10.1371/journal.pmed.1000335

Gellad, W. F., Flynn, K. E., & Alexander, G. C. (2015). Evaluation of flibanserin: Science and advocacy at the FDA. *Journal of the American Medical Association*, *314*(9), 869–870. https://doi.org/10.1001/jama.2015.8405

Goldstein, S. W. (2009). My turn . . . Finally. *Journal of Sexual Medicine*, *6*(2), 301–302. https://doi.org/10.1111/j.1743-6109.2008.01147.x

Graham, C. A., Boynton, P. M., & Gould, K. (2017). Women's sexual desire: Challenging narratives of "dysfunction.". *European Psychologist*, *22*(1), 27–38. https://doi.org/10.1027/1016-9040/a000282

Graham, C. A., Brotto, L. A., & Zucker, K. J. (2014). Response to Balon and Clayton (2014): Female sexual interest/arousal disorder is a diagnosis more on firm ground than thin air. *Archives of Sexual Behavior*, *43*(7), 1231–1234. https://doi.org/10.1007/s10508-013-0248-0

Hart, B., Lundh, A., & Bero, L. (2012). Effect of reporting bias on meta-analyses of drug trials: Reanalysis of meta-analyses. *BMJ*, *344*(jan03 1), d7202. https://doi.org/10.1136/bmj.d7202

Healy, D. (2006). The latest mania: Selling bipolar disorder. *PLoS Medicine*, *3*(4), e185. https://doi.org/10.1371/journal.pmed.0030185

Healy, D., & Cattell, D. (2003). Interface between authorship, industry and science in the domain of therapeutics. *British Journal of Psychiatry*, *183*(1), 22–27. https://doi.org/10.1192/bjp.183.1.22

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ (British Medical Journal)*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hogenmiller, A., Aless, H., & Fugh-Berman, A. (2017, June 14). *The score is even*. The Hastings Center. Retrieved from The Hastings Center website: https://www.thehastingscenter.org/the-score-is-even/

Horwitz, A. V., & Wakefield, J. C. (2007). *The loss of sadness: How psychiatry transformed normal sorrow into depressive disorder*. Oxford University Press.

Hughes, S., Cohen, D., & Jaggi, R. (2014). Differences in reporting serious adverse events in industry sponsored clinical trial registries and journal articles on antidepressant and antipsychotic drugs: A cross-sectional study. *BMJ Open*, *4*(7), e005535. https://doi.org/10.1136/bmjopen-2014-005535

Hyman, S. E. (2010). The diagnosis of mental disorders: The problem of reification. *Annual Review of Clinical Psychology*, *6*(1), 155–179. https://doi.org/10.1146/annurev.clinpsy.3.022806.091532

Jaspers, L., Feys, F., Bramer, W. M., Franco, O. H., Leusink, P., & Laan, E. T. M. (2016). Efficacy and safety of flibanserin for the treatment of hypoactive sexual desire disorder in women: A systematic review and meta-analysis. *JAMA Internal Medicine*, *176*(4), 453. https://doi.org/10.1001/jamainternmed.2015.8565

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Jureidini, J. N., Amsterdam, J. D., & McHenry, L. B. (2016). The citalopram CIT-MD-18 pediatric depression trial: Deconstruction of medical ghostwriting, data mischaracterisation and academic malfeasance. *The International Journal of Risk & Safety in Medicine*, *28*(1), 33–43. https://doi.org/10.3233/JRS-160671

Jureidini, J. N., & McHenry, L. B. (2020). *The illusion of evidence-based medicine: Exposing the crisis of credibility in clinical research*. Wakefield Press.

Jureidini, J. N., McHenry, L. B., & Mansfield, P. R. (2008). Clinical trials and drug promotion: Selective reporting of study 329. *International Journal of Risk & Safety in Medicine*, *20*(1–2), 73–81. https://doi.org/10.3233/JRS-2008-0426

Jutel, A. (2010). Framing disease: The example of female hypoactive sexual desire disorder. *Social Science & Medicine*, *70*(7), 1084–1090. https://doi.org/10.1016/j.socscimed.2009.11.040

Kaplan, R. M., Irvin, V. L., & Garattini, S. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over Time. *Plos One*, *10*(8), e0132382. https://doi.org/10.1371/journal.pone.0132382

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Kingsberg, S. A., Clayton, A. H., Portman, D., Williams, L. A., Krop, J., Jordan, R., Lucas, J., & Simon, J. A. (2019). Bremelanotide for the treatment of hypoactive sexual desire disorder: Two randomized phase 3 trials. *Obstetrics and Gynecology*, *134*(5), 899–908. https://doi.org/10.1097/AOG.0000000000003500

Kirsch, I., & Moncrieff, J. (2007). Clinical trials and the response rate illusion. *Contemporary Clinical Trials*, *28*(4), 348–351. https://doi.org/10.1016/j.cct.2006.10.012

Kraemer, H. C., & Robinson, T. N. (2005). Are certain multicenter randomized clinical trial structures misleading clinical and policy decisions? *Contemporary Clinical Trials*, *26*(5), 518–529. https://doi.org/10.1016/j.cct.2005.05.002

Lane, C. (2008). *Shyness: How normal behavior became a sickness*. Yale University Press.

Laumann, E. O., Paik, A., & Rosen, R. C. (1999). Sexual dysfunction in the United States: Prevalence and predictors. *JAMA*, *281*(6), 537–544. https://doi.org/10.1001/jama.281.6.537

Le Noury, J., Nardo, J. M., Healy, D., Jureidini, J., Raven, M., Tufanaru, C., & Abi-Jaoude, E. (2015). Restoring study 329: Efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ*, *351*(8025), h4320. https://doi.org/10.1136/bmj.h4320

Logdberg, L. (2011). Being the ghost in the machine: A medical ghostwriter's personal view. *PLOS Medicine*, *8*(8), e1001071. https://doi.org/10.1371/journal.pmed.1001071

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19. https://doi.org/10.1037/1082-989X.7.1.19

Masouleh, S. K., Eickhoff, S. B., Hoffstaedter, F., Genon, S., & Initiative, A. D. N. (2019). Empirical examination of the replicability of associations between brain structure and psychological variables. *Elife*, *8*, e43464. https://doi.org/10.7554/eLife.43464

Matheson, A. (2016a). Ghostwriting: The importance of definition and its place in contemporary drug marketing. *BMJ*, *354*(8071). https://doi.org/10.1136/bmj.i4578

Matheson, A. (2016b). The disposable author: How pharmaceutical marketing is embraced within medicine's scholarly literature. *Hastings Center Report*, *46*(4), 31–37. https://doi.org/10.1002/hast.576

Mathieu, S., Boutron, I., Moher, D., Altman, D. G., & Ravaud, P. (2009). Comparison of registered and published primary outcomes in randomized controlled trials. *Journal of the American Medical Association*, *302*(9), 977–984. https://doi.org/10.1001/jama.2009.1242

Mathieu, S., Chan, A.-W., Ravaud, P., & Smalheiser, N. R. (2013). Use of trial register information during the peer review process. *PLoS ONE*, *8*(4), e59910. https://doi.org/10.1371/journal.pone.0059910

Mayo-Wilson, E., Fusco, N., Li, T., Hong, H., Canner, J. K., Dickersin, K., Cowley, T., Doshi, P., Ehmsen, J., Gresham, G., Guo, N., Haythornthwaite, J., Heyward, J., Pham, D., Payne, J., Rosman, L., Stuart, E., Suarez-Cuervo, C., Tolbert, E., Vedula, S., & Bertizzolo, L. (2019a). Harms are assessed inconsistently and reported inadequately Part 1: Systematic adverse events. *Journal of Clinical Epidemiology*, *113*(3), 20–27. https://doi.org/10.1016/j.jclinepi.2019.04.022

Mayo-Wilson, E., Fusco, N., Li, T., Hong, H., Canner, J. K., Dickersin, K., Cowley, T., Doshi, P., Ehmsen, J., Gresham, G., Guo, N., Haythornthwaite, J., Heyward, J., Pham, D., Payne, J., Rosman, L., Stuart, E., Suarez-Cuervo, C., Tolbert, E., Vedula, S., & Bertizzolo, L. (2019b). Harms are assessed inconsistently and reported inadequately Part 2: Nonsystematic adverse events. *Journal of Clinical Epidemiology*, *113*(3), 11–19. https://doi.org/10.1016/j.jclinepi.2019.04.020

McHenry, L. B., & Amsterdam, J. D. (2019). The paroxetine 352 bipolar study revisited: Deconstruction of corporate and academic misconduct. *Journal of Scientific Practice and Integrity*, *1*(1). https://doi.org/10.35122/jospi.2019.958452

McHenry, L. B., & Jureidini, J. N. (2008). Industry-sponsored ghostwriting in clinical trial reporting: A case study. *Accountability In Research*, *15*(3), 152–167. https://doi.org/10.1080/08989620802194384

Meixel, A., Yanchar, E., & Fugh-Berman, A. (2015). Hypoactive sexual desire disorder: Inventing a disease to sell low libido. *Journal of Medical Ethics*, *41*(10), 859–862. https://doi.org/10.1136/medethics-2014-102596

Merriam-Webster. (2020, January 20). *Definition of ghostwriting*. Merriam-Webster. https://www.merriam-webster.com/dictionary/ghostwriting

Mills, J. L. (1993). Data torturing. *The New England Journal of Medicine*, *329*(16), 1196–1199. https://doi.org/10.1056/NEJM199310143291613

Mitchell, K. R., Jones, K. G., Wellings, K., Johnson, A. M., Graham, C. A., Datta, J., Bancroft, J., Sonnenberg, P., Macdowall, W., Field, N., Mercer, C. H., & Copas, A. J. (2016). Estimating the prevalence of sexual function problems: The impact of morbidity criteria. *Journal of Sex Research*, *53*(8), 955–967. https://doi.org/10.1080/00224499.2015.1089214

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gotzsche, P., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*(mar23 1), c869. https://doi.org/10.1136/bmj.c869

Moynihan, R. (2003). The making of a disease: Female sexual dysfunction. *BMJ*, *326*(7379), 45–47. https://doi.org/10.1136/bmj.326.7379.45

Neijenhuijs, K. I., Hooghiemstra, N., Holtmaat, K., Aaronson, N. K., Groenvold, M., Holzner, B., Terwee, C. B., Cuijpers, P., & Verdonck-de Leeuw, I. M. (2019). The Female Sexual Function Index (FSFI)—A systematic review of measurement properties. *The Journal of Sexual Medicine*, *16*(5), 640–660. https://doi.org/10.1016/j.jsxm.2019.03.001

Obstetrics & Gynecology. (2019). *RE: Manuscript number ONG-19-803*. Lippincott Williams & Wilkins. Retrieved from Your Submission ONG-19-80e website: https://cdn-links.lww.com/permalink/aog/b/aog_134_5_2019_09_12_kingsberg_19-803_sdc2.pdf

Obstetrics & Gynecology. (2020). *Instructions for authors: Obstetrics & gynecology*. Lippincott Williams & Wilkins. https://journals.lww.com/greenjournal/Pages/instructionsforauthors.aspx

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), 6251. https://doi.org/10.1126/science.aac4716

Palatin Technologies. (2014). *A placebo-controlled, randomized, parallel group, dose-finding trial to evaluate the efficacy and safety of subcutaneously administered bremelanotide in premenopausal women with FSAD (Female Sexual Arousal Disorder) and/or HSDD (Hypoactive Sexual Desire Disorder)* (Clinical Trial Registration No. NCT01382719). clinicaltrials.gov. Retrieved from clinicaltrials.gov website: https://clinicaltrials.gov/ct2/show/NCT01382719

Paris, J. (2009). The bipolar spectrum: A critical perspective. *Harvard Review of Psychiatry*, *17*(3), 206–213. https://doi.org/10.1080/10673220902979888

Paris, J. (2015). *Overdiagnosis in psychiatry: How modern psychiatry lost its way while creating a diagnosis for almost all of life's misfortunes*. Oxford University Press.

Phase Five Publications. (n.d.). *This is grey*. Phase Five Publications. https://www.phase-five.com/sites/phasefivecom/files/pdfs/Doing_It_Right.pdf

Phase Five Communications. (2020). *Phase five communications*. Phase Five Communications. Retrieved from Phase Five | Home website: https://www.phase-five.com/

Revicki, D. A., Clayton, A. H., Stouch, B. C., Portman, D. J., Kingsberg, S. A., DeRogatis, L. R., & Jordan, R. (2018, February 20).

Developing clinically meaningful responder thresholds for primary endpoints for clinical trials in premenopausal women with hypoactive sexual desire disorder. Presented at the International Society for CNS Trials and Methodology, Washington, DC.

Roest, A. M., de Jonge, P., Williams, C. D., de Vries, Y. A., Schoevers, R. A., & Turner, E. H. (2015). Reporting bias in clinical trials investigating the efficacy of second-generation antidepressants in the treatment of anxiety disorders: A report of 2 meta-analyses. *JAMA Psychiatry*, *72*(5), 500–510. https://doi.org/10.1001/jamapsychiatry.2015.15

Rosen, R. (2000). The Female Sexual Function Index (FSFI): A multidimensional self-report instrument for the assessment of female sexual function. *Journal of Sex & Marital Therapy*, *26*(2), 191–208. https://doi.org/10.1080/009262300278597

Ross, J. S., Hill, K. P., Egilman, D. S., & Krumholz, H. M. (2008). Guest authorship and ghostwriting in publications related to rofecoxib: A case study of industry documents from rofecoxib litigation. *JAMA*, *299*(15), 1800–1812. https://doi.org/10.1001/jama.299.15.1800

Ross, J. S., Mulvey, G. K., Hines, E. M., Nissen, S. E., Krumholz, H. M., & Sim, I. (2009). Trial publication after registration in ClinicalTrials.Gov: A cross-sectional analysis. *PLoS Medicine*, *6*(9), e1000144. https://doi.org/10.1371/journal.pmed.1000144

Sakaluk, J. K., & Graham, C. A. (2018). Promoting transparent reporting of conflicts of interests and statistical analyses at The Journal of Sex Research. *Journal of Sex Research*, *55*(1), 1–6. https://doi.org/10.1080/00224499.2017.1395387

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *PLoS Medicine*, *7*(3), e1000251. https://doi.org/10.1371/journal.pmed.1000251

Schwarz, A. (2016). *ADHD nation*. Scribner.

Segal, J. Z. (2015). The rhetoric of female sexual dysfunction: Faux feminism and the FDA. *CMAJ*, *187*(12), 915–916. https://doi.org/10.1503/cmaj.150363

Segal, J. Z. (2018). Sex, drugs, and rhetoric: The case of flibanserin for 'female sexual dysfunction.'. *Social Studies of Science*, *48*(4), 459–482. https://doi.org/10.1177/0306312718778802

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Sismondo, S. (2007). Ghost management: How much of the medical literature is shaped behind the scenes by the pharmaceutical industry? *PLoS Medicine*, *4*(9), e286–e286. https://doi.org/10.1371/journal.pmed.0040286

Sismondo, S. (2018). *Ghost-managed medicine: Big pharma's invisible hands*. Mattering Press.

Sismondo, S., & Nicholson, S. H. (2009). Publication planning 101. *Journal of Pharmacy & Pharmaceutical Sciences*, *12*(3), 273–279. https://doi.org/10.18433/J3WW2R

Spielmans, G. I. (2009). The promotion of olanzapine in primary care: An examination of internal industry documents. *Social Science & Medicine*, *69*(1), 14–20. https://doi.org/10.1016/j.socscimed.2009.05.001

Spielmans, G. I., Berman, M. I., Linardatos, E., Rosenlicht, N. Z., Perry, A., & Tsai, A. C. (2013). Adjunctive atypical antipsychotics for major depressive disorder: A meta-analysis of depression, quality of life, and safety outcomes. *PLoS Medicine*, *10*(3), e1001403. https://doi.org/10.1371/journal.pmed.1001403

Spielmans, G. I., & Parry, P. I. (2010). From evidence-based medicine to marketing-based medicine: Evidence from internal industry documents. *Journal of Bioethical Inquiry*, *7*(1), 13–29. https://doi.org/10.1007/s11673-010-9208-8

Szucs, D., Ioannidis, J. P. A., & Wagenmakers, E.-J. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, *15*(3), e2000797. https://doi.org/10.1371/journal.pbio.2000797

Tavernise, S., & Pollack, A. (2015, June 13). Aid to women, or bottom lne? Advocates split on libido pill. *The New York Times*. https://www.nytimes.com/2015/06/14/us/aid-to-women-or-bottom-line-advocates-split-on-libido-pill.html

Tiefer, L. (2006). Female sexual dysfunction: A case study of disease mongering and activist resistance. *PLOS Medicine*, *3*(4), e178. https://doi.org/10.1371/journal.pmed.0030178

Turner, E. H. (2013). Publication bias, with a focus on psychiatry: Causes and solutions. *CNS Drugs*, *27*(6), 457–468. https://doi.org/10.1007/s40263-013-0067-9

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine*, *358*(3), 252–260. https://doi.org/10.1056/NEJMsa065779

United States Food and Drug Administration. (2019). United States Food and Drug Administration. Retrieved from New Drug Application Multi-Disciplinary Review and Evaluation Standard 210557: Vyleesi/Bremelanotide website: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2019/210557Orig1s000TOC.cfm