PUBLIC VERSION

Before the
UNITED STATES COPYRIGHT ROYALTY JUDGES
THE LIBRARY OF CONGRESS
Washington, D.C.

In the Matter of	)	
DETERMINATION OF RATES AND TERMS FOR DIGITAL PERFORMANCE IN SOUND RECORDINGS AND EPHEMERAL RECORDINGS (WEB IV)	)	**Docket No. 14-CRB-0001-WR**

WRITTEN DIRECT TESTIMONY OF STEPHAN MCBRIDE

(On behalf of Pandora Media, Inc.)

Introduction

1. My name is Stephan McBride. I am a researcher on the Science Team at Pandora Media, Inc. ("Pandora"), Pandora's internal group of researchers tasked with designing studies and performing analyses to improve the listening experience of the service. My title is Senior Scientist, Economics, a position I have held since joining the Company in April 2014.

2. I graduated from Stanford University with a Ph.D. in Economics in 2008. I also hold graduate degrees from Yale University and Queen's University. Prior to joining Pandora, I worked as a professional economist consulting companies on pricing, valuation, and intellectual property matters. I designed and analyzed numerous economics studies, including experimental studies, during this period. Including my experience during graduate studies, I have over a decade of experience performing and analyzing economic studies.

3. The primary focus of my role at Pandora is designing, executing, and analyzing studies and experiments that measure how Pandora's service (and changes to features of the service) impacts listener retention, pricing, and the broader music industry. In the performance of my duties, I regularly work with my Science Team colleagues, Pandora software engineers,

and business managers to create and implement analyses that could produce guidance or answers to questions of business relevance to Pandora. For example, we recently investigated the effect on the hours listened and retention of listeners of recommending to listeners additional stations they may wish to create. Other experiments have addressed topics such as the effect of different compensation packages for Pandora sales personnel, the effectiveness of music recommendation based on computer-analysis of the music signal, the effect of push notifications, and effectiveness of targeted ads for Pandora One.

4. I submit this testimony to describe for the Copyright Royalty Judges the research methods used by the Science Team in general, and the specific application of these research methods to the design and analysis of two studies created to assess the impact of changes to the core Pandora music recommendation algorithms: what I refer to below as the "Steering Experiments" and the "Music Sales Experiments." Although I undertook these two studies at the request of counsel in relation to this proceeding, they are materially equivalent to the types of studies that I and other members of the Science Team regularly perform in the course of our roles at Pandora.[1]

Scientific Research at Pandora

5. The Pandora Science Team performs rigorous research and analyses to measure the effectiveness of Pandora features to improve the listening experience on Pandora. The Science Team is comprised of 15 individuals, 13 of whom hold doctorate degrees in computer science, engineering, statistics, or economics from leading academic institutions, and who have considerable substantive expertise in areas related to Pandora's core mission.

[1] For example, the Steering Experiments discussed here follow similar experiments conducted in 2013 at the request of management to inform important business decisions regarding entering into direct licenses.

6. The Science Team designs and executes analyses using the vast data available at Pandora. We employ standard methods used by statisticians, computer scientists, and economists working in academia and at other technology companies. These methods can include traditional statistical methods such as linear regression or logistic regression, as well as machine learning approaches to analyze large quantities of data (a topic I address in more detail below).

7. The Science Team has primary responsibility for designing and analyzing controlled experiments, the most rigorous approach to data analysis. Controlled experiments are comparisons between groups of listeners, one of which randomly receives a manipulated experience (the "treated" group) and the other of which randomly receives the standard Pandora experience (the "control" group). When experiments are (1) randomized, (2) controlled, and (3) blind,[2] they represent the "gold standard"[3] for determining the causal impact of the manipulated experience. All Pandora experiments meet these gold-standard experimental requirements, ensuring that comparisons between the treated and control groups provide the most rigorous estimate of the effect of the changed experience that can be identified.[4]

8. Pandora has set up a framework for conducting experiments in which a Pandora Scientist or Engineer intentionally changes one aspect of the Pandora experience for a sample

[2] "Randomized" means assignment of listeners to treatment is ultimately random, in contrast to deterministic assignment by the researcher. "Controlled" means the primary outcome is a comparison between those receiving the exposure and those not to account for the 'placebo effect'. "Blind" means experimental subjects are unaware of their assignment to the treatment or control. In digital experimentation, blinding goes further by masking subjects from their participation in the experiment. Random assignment is critical to a well-designed experiment to remove researcher bias, to balance other factors that could affect the result (confounders), to ensure valid causal inference (whether A did or did not cause B, with what level of statistical confidence), and, most importantly, to ensure the generalizability of results.

[3] Such experiments, for example, are commonly used by the FDA to test adverse drug impacts, as well by courts. *See* Green, Michael D., D. Michal Freedman, and Leon Gordis, "Reference Guide on Epidemiology", in *Reference Manual on Scientific Evidence, 3rd Edition*, p. 555.

[4] Currently, there are over 65 experiments operated by Pandora researchers.

group of listeners and then compares the effects to groups of listeners who did not experience the change. We refer to this approach as "A/B testing," where A is the control group and B is the group of listeners that receives the manipulated experience.

Pandora constructed its A/B framework to support controlled experimentation that takes advantage of Pandora's vast scale (76+ million monthly active users with over 800 million spins of music daily) and two-way communication with its users. Unlike terrestrial radio, we receive detailed information about our users' listening habits: listening time, retention, "thumbs up" and "thumbs down," skips, station creation, subscription patterns, and the like. This has enabled, for the first time in a radio environment, the use of controlled experimentation to investigate, with sufficient power and reliability, how changes in programming and service features (among other things) affect listening behavior.

9. Such experimentation, when designed properly and executed correctly, is superior to observational studies because it directly addresses the research question at hand without additional (and untestable) assumptions. The external manipulation to create A/B comparison groups balances differences between the groups so that the comparison between treated and control groups reflects only the difference caused by the manipulation, and assures that other factors did not in fact cause the observed difference. Observational studies, by comparison, attempt to translate an association (e.g., we see more purchases where there are more ads) into a cause (*because* there were more ads, and *not because* of any other factor, we saw more purchases).[5]

[5] Consider a concrete example relevant to Pandora. Because we observe that listeners who create more personal Pandora stations listen more, we might ask whether Pandora should therefore encourage its users to create more stations (for example, by sending them emails highlighting that feature). It might be the case, however, that causation runs the opposite way: when users listen more frequently, they create more

10. This testing ability and the level of rigor with which such tests are conducted also allows Pandora to assess and improve the personalized Internet radio service we provide our listeners, the targeted advertising we use to assist our advertisers, and the revenue we provide to music creators – artists and their labels.

11. The distinction between research on Pandora using experimental methods and previous research on radio listening cannot be overstated. Consider a likely familiar problem: did service *X* (terrestrial radio/satellite radio/Pandora) increase exposure to a new artist ("New Act")? For researchers analyzing the impact of, say, terrestrial radio, the two possible approaches would be to survey listeners ("Do you know New Act? And if so, where did you first hear about New Act?") or to compare listening of New Act across regions (arguing that increased listening was primarily attributable to their specific radio exposure in that region). Neither approach, however, can control for other reasons – differential exposure to social media across interviewees/regions, recall biases, regional genre preference, or even the "home town" effect of New Act band members – that could explain differential familiarity. Accordingly, results from these analyses are typically anecdotal and not generalizable. In contrast, random assignment of exposure to New Act on Pandora across listeners (making the band's music available to some listeners but not others) accounts for these differences so that tracking, say, stations created for New Act or total spins for New Act (treatment v. control) provides a direct measure of the impact of Pandora's increased exposure.

The Steering Experiments

12. For the Steering Experiments, we were attempting to answer the question of whether increases or decreases in performances of sound recordings owned by a particular record

---

stations. Only a controlled experiment in which Pandora encourages station creation for a treatment (B) group and compares resulting listening behavior to the control group (A) will answer the question.

company would have a measurable impact on a key listener metric, specifically average hours listened per registered user.[6] Secondarily, these experiments were also designed to test/substantiate Pandora's engineering capability to precisely manipulate the share of music played according to various criteria (here, the record company that owns the recordings).

13. To answer these questions, my colleagues and I intentionally manipulated the share of music played on the service for test groups of listeners based on the companies that own the sound recordings. Three ownership groups were considered in these experiments: Universal Music Group ("UMG"), Sony Music ("Sony"), and Warner Music Group ("WMG"); all other sound recordings were classified as "Other." We conducted the experiments at the request of counsel and with the input of Professor Carl Shapiro, who provided detailed instructions ("Shapiro Directions") as described below.[7] Although undertaken pursuant to Professor Shapiro's direction, these experiments closely match previous experiments that the Science Team ran between summer 2013 and January 2014 for business investigations in which the core music recommendation algorithms were manipulated to increase/decrease the spin share of music based on the ownership of the sound recording.

14. I was deeply involved in the design, execution, and preparation of results for these experiments. To the extent certain tasks related to the experiments were handled by my colleagues, I am well aware of those aspects and can speak to them. I provide an overview of the experimental design and analysis here; Technical Appendix A includes additional detail.

[6] This is the most commonly studied measure we use for assessing changes to the core Pandora music recommendation algorithms. Almost all listening on Pandora is by users who are registered.

[7] These instructions were followed in all material respects. The only exception: the experiments ended on September 3, 2014, two days later than directed, to ensure the final period had a full week's worth of data. A member of Professor Shapiro's team approved this extension prior to the originally scheduled end date.

Study Details

15. The Steering Experiments consisted of a group of 12 experiments, each defined by a combination of a target ownership group (UMG/Sony/WMG) and a target deflection in share of spins (treatment group) as compared to spins that would occur according to the standard Pandora music recommendation results (control group). The requested spin share deflections (the "steering") were: -30%, -15%, +15%, and +30% for each of the three ownership groups manipulated. These percentage deflections comprise the experimental manipulations. These experiments started on June 4, 2014 and ended on September 3, 2014 (13 weeks).

16. The Steering Experiments operated through the above-described A/B Framework. The experimental subjects were all Pandora listeners, each of whom was randomly assigned to one of the 12 treatment groups, to the single control group, or were included in the portion of listeners excluded from all experiments. The treatment groups for UMG each had 5% of listeners, the treatment groups for Sony each had 7% of listeners, and the treatment groups for WMG each had 8% of listeners; the control group had 10% of listeners.[8]

17. The Steering Experiments were implemented by manipulating the probability that any song from the target music group would play on Pandora using a single manipulation factor. The single manipulation factor applied to all situations – ignoring all information about listener, sound recording, and station – and thus was a "naïve" manipulation, producing the largest reasonable estimate of listening impact; a targeted approach that steered based on knowledge of

[8] To increase statistical reliability, the treatment group sizes varied across target music group to account for differences in share of music played on Pandora, having larger sample sizes for the smaller major music groups.

user characteristics or listening context would almost certainly yield a smaller listening impact.[9] The manipulation factor was calculated nightly for each experiment based on the achieved difference in spin share of the target; the new factor applied to the subsequent day with daily updates in the factor until a value was achieved that yielded the desired level of steering.

18. The Steering Experiments used music ownership information previously collected for Pandora business purposes, with ownership combined into the three target music groups: UMG, Sony, and WMG. Technical Appendix A provides additional detail. At all times during the experiments, the steering used the most current list of music ownership, and applied the most current list to assess steering accuracy for all previous weeks.

Analysis

19. The Shapiro Directions set forth that the experiments be analyzed using the standard listener metric that Pandora itself tracks when running experiments, namely, average hours listened per registered listener. The A/B Framework calculates this metric daily for 7-day periods, storing the information in databases. I personally queried these databases and provided the extracted results to Professor Shapiro's team. I also calculated average hours listened per registered listener for the entire experimentation period. To confirm the experimental manipulation, I calculated the spin share for the target music group for each experiment and provided these to Professor Shapiro's team.

[9] For example, we might decide to reduce plays solely of "album cuts" from a given Sony album, but not the lead singles; or, we might not deflect Sony tracks on listener stations seeded with Mariah Carey, a prominent Sony recording artist. Additional factors that might be elements of the steering rules: decreased steering with duration of listening on a particular station, or greater steering for listeners who use Pandora less frequently, or number of thumbed up songs on a station (with greater steering for stations that have many available songs).

Results

20. **Figure 1** presents a graphic illustration of the experimental manipulation. The figure presents the change in spin share for the target ownership group for each of the 12 treatment groups versus the control; the actual manipulations closely approximate the target steering. Two outliers are however evident: the first week during which the manipulation factor was adjusted most significantly on a daily basis (to obtain the desired increase/decrease in record company spin share), and the week including August 18, 19, and 20. For those days, a malfunction in the A/B Framework prevented the operation of the steering manipulation. Those outliers aside, this figure demonstrates that Pandora's playlist algorithm can precisely manipulate the share of music played according to a specified criterion, as the change in the actual spin share closely matches the target spin share.

21. The experimental manipulations had very minor, and in most cases, statistically insignificant impacts on listening on Pandora; that is, Pandora is able to steer both toward and away from music of the three investigated music groups with minimal or no effect on the Pandora listening experience. **Figure 2** presents the change in average hours listened, weekly, between each of the 12 treatment groups and the control.[10] Notably, these results support that

[redacted]

[10] The figure presents a grid, with rows for the three music group targets (UMG, Sony, WMG), and the columns for the four steering targets (-30%, -15%, +15%, +30%). The effect estimate each week is the dot, with its 95% confidence interval the attached vertical line.

[11] Results were statistically insignificant for WMG.

As shown in **Table 1**, the average effects on listening hours for the entire experimental period were even smaller (and less statistically significant).[12]

22. I reiterate that these small listening impacts are likely larger than would occur if the steering had been more targeted – i.e., if we reduced or increased particular plays of the record company repertoire differently depending on context (listener characteristics, sound recording characteristics, and station characteristics).

Music Sales Experiments

23. The Music Sales Experiments ("MSEs") were controlled experiments designed to test whether performances of sound recordings on Pandora have a positive or negative impact on sales of those sound recordings. As described below, we intentionally manipulated the availability of certain recordings on Pandora in certain locations and measured the effect on sales of the same music. This work was performed at the request of counsel in this matter, but is related to business inquiries performed by Pandora researchers to examine Pandora's effect on the music industry. Consistent with Pandora practices, the Music Sales Experiments were broad collaborations. I participated in all facets of the design, execution, data merging, quality checking, and data analyses. Two of my Pandora colleagues were instrumental contributors. Jonah Liebert, an experienced Pandora engineer, constructed and tested the code that turned off music for listeners based on geography. Oliver Bembom, another Pandora Senior Scientist who holds a Ph.D. in Statistics from UC Berkeley, and who has worked extensively in experimental design and analysis, performed much of the data analysis. I am aware of Jonah and Oliver's work on the experiments and can speak to it here.

[12] **Table 1** presents the average change in listening hours treatment v. control for the 12 experiments for the entire experimental period (13 weeks), and for the entire experimental period excluding the week during which the system malfunction occurred (week starting August 14, 2014).

Study Details

24. The Music Sales Experiments are a group of research experiments in which we prevented the performance, for all listeners in randomly selected regions, of either (i) recordings new to Pandora ("New MSEs"); or (ii) catalog recordings long spinning on Pandora ("Catalog MSEs"). The experimental manipulation was to *not* spin the referenced sound recordings on Pandora during the experimental period. In contrast to the Steering Experiments, the manipulation was not a change to the core Pandora music recommendation algorithms, but rather a complete disabling of those recordings in the randomly selected geographic regions. For listeners not treated with the experimental manipulation (the control group), music played according to standard processes by which Pandora creates playlists.

25. The experimental manipulation – not spinning on Pandora during the experimental period – occurred at the region level in this experiment rather than the listener level. (Pandora lacks music sales data at the listener level.) To match the experimental manipulation to the available sales data, we segmented listeners according to the most granular region for which we could access weekly music sales data.

26. **Table 2** presents a summary of the New MSEs and the Catalog MSEs. All experiments occurred between June and September 2014, and all were 8 weeks in duration, except for the second Catalog MSEs.[13] The New MSEs manipulated all tracks on albums that were new to Pandora, while those in the Catalog MSEs manipulated specific catalog songs.[14]

[13] The earliest New MSEs started on June 10, 2014; the last New MSEs ended September 3, 2014. The first Catalog MSEs started on July 16, 2014, and the second Catalog MSEs (explained in detail below) started on August 12, 2014. All Catalog MSEs ended September 9, 2014.

[14] The Catalog MSEs used particular tracks because catalog recordings are often available on multiple albums, requiring an infeasible escalation of the research if all such albums were included.

There were 1,215 experiments (814 New MSEs; 401 Catalog MSEs) with at least one unit sold during the experimentation period.

27. [redacted][15] When including SoundScan's "extended DMAs" (DMAs not in the top 100), and "sub DMAs" (subdivisions of the five largest DMAs), SoundScan tracks sales for 228 mutually-exclusive US regions.[16] All Pandora listeners with a zip code mapping to one of these DMA regions were included in the experiments.[17]

28. The Music Sales Experiments were not operated through the A/B Framework because that framework lacks the capacity to randomize listeners based on location of residence. Accordingly, my colleagues and I created an algorithm that randomized listeners to treatment and control groups based on their DMAs. We performed tests on the randomization to confirm the randomness of assignment. In addition, rather than apply a single randomization to all sound recordings in the experiments, we randomized the experimental manipulation for *each* sound

[15] *See*, http://www.nielsen.com/intl-campaigns/us/dma-maps.html, accessed August 2014.

[16] SoundScan tracks sales for 230 US regions. However, Pandora internal data combine two of the SoundScan DMAs (Tuscaloosa and Anniston, AL) into a single region. We discovered this difference after creating the randomization for use in the first MSEs and submitting for use in the Pandora music system. Accordingly, listeners in these two DMAs were jointly randomized to treatment or control, a change in experimental design. We removed all listeners in these DMAs from the experiments; producing an effective 228 US regions.

[17] The vast majority of Pandora listeners in the US were included in the experiments. As of May 2014, 97.3% of all listeners had a Zip Code and 98.73% of these listeners were resident in one of the 228 SoundScan DMAs for which we have data.

recording tested (that is to say, the geographic locations of listeners in the control and treatment groups varied for the different manipulated sound recordings). This approach helped to ensure that no predictable purchasing patterns, such as genre-preference differences in particular regions, would affect the results, making them more robust. In short, the experimental design of the Music Sales Experiments was randomized, controlled, and blind, consistent with everything in the A/B Framework we normally employ.

29. The study looked at aggregate music sales in each DMA for each track/album as reported by SoundScan for the experimentation period. SoundScan tracks unit sales (physical and digital) for weeks starting on Monday, along with location of sale for the majority of music sold in the U.S.[18] We decided to use SoundScan to measure sales as this is the industry standard – for example, it is the data used by the recording industry itself to measure and report sales.

30. [redacted] To prepare unit sales for each experiment in each DMA, we converted single sales to album sales using SoundScan's standard definition of the "Track-equivalent album"

[redacted]

("TEA"): one album equals 10 singles. TEA sales are the units we measure in the New MSEs.[21] [redacted] Track sales are the units we measure in the Catalog MSEs.

31. For the New MSEs, we pre-selected a week for the experiments to start, and identified all music that would newly spin on Pandora that week for inclusion in the experiments. We had no advance information about the music that would be released that week when making the determination for the start week; we also did not inform any Pandora personnel involved in new music ingestion about the impending experiment. We know of no reason to deem the initial week exceptional.[22] We later added additional rounds of New MSEs, applying the same "all comers" inclusion rules to subsequent weeks. At no point did we have advanced notice of what music was going to launch on Pandora on the selected weeks.

32. We applied two exclusion criteria to the "all comers" rule in the New MSEs to focus on albums (our unit of manipulation). First, we excluded follow-on singles – a new single from an unreleased album that had a previous single already spinning; the previous single was expected to comprise the bulk of spins. Second, in some instances, sound recordings that were on albums new to Pandora were not in fact actually "new." The same sound recording was included on another album already spinning on Pandora (for example, when a greatest hits album

[21] For a very small proportion of observations, we observed TEA sales smaller than zero. [redacted] Such negative TEA sales were set to zero before analyzing the data.

[22] We devised rules for inclusion in the MSEs to exclude our own or other Pandora personnel from selecting the music, and thus introducing selection concerns. The rules effectively tied our hands.

is released that is just a compilation of previously released sound recordings). To ensure that the New MSEs in fact only contained music new to Pandora, we excluded these tracks.[23]

33. For the Catalog MSEs, as with the New MSEs, we used inclusion rules that reduced researcher selection to guard against the introduction of bias. For each of the two Catalog MSEs rounds, we randomly selected 200 songs that met our weekly spin and music sale specifications from the *Rolling Stone Top 500 Songs* (from April 2011) and the *Pitchfork 500* (based on songs released during the 1977-2006 period) – both lists of significant songs.[24]

34. These lists represent music deemed significant by music critics. We selected these lists so that we would be manipulating music that is broadly familiar, and that can be quickly accessed on genre stations on Pandora, among numerous other sources. These characteristics should, if anything, decrease the measured promotional effect of Pandora: because the music is familiar, one would assume it is harder for Pandora to introduce / re-introduce the music to listeners, thereby reducing the promotional effect. For these reasons, we deemed these lists a superior population to sample from to assess the Pandora effect on sales.

[23] When a recording new to Pandora was a new version of a song by the same artist that was already available on Pandora, we made sure to include all versions of that song by that artist in the experiment. We did this so as to prevent the earlier track, an extremely close musicological substitute, from being played by the standard Pandora music recommendation algorithms for listeners precluded from hearing the manipulated track.

[24] Inclusion criteria were minimum use on Pandora (spins), and minimum sales. See Technical Appendix B for additional detail. We excluded songs from consideration that did not meet inclusion criteria; we also excluded songs that we could not identify in either Pandora or SoundScan data. The minimum sales requirement explains why we focused on songs rather than albums for the catalog experiments: catalog albums – even those in the *Rolling Stone Top 500 Albums* list, for example – tend to have very limited sales, and sales of classic songs tend to be dispersed across multiple albums (notably, greatest hits albums). It was not feasible to remove and manipulate every album with any song from an album on the *Rolling Stone Top 500 Albums* list.

35. All MSEs were ended at their scheduled time, after eight-week experimentation periods for all experiments other than those in the second round of Catalog MSEs, for which, due to data limitations, we used a four-week experimentation period.

Analysis

36. The variability of music sales is well known, with some sound recordings selling tens or hundreds of thousands of tracks / albums per week, and others receiving no or negligible sales. Some of these differences in unit sales are predictable, for example based on the popularity of the artist as measured by sales of a recent album, or by spins on Pandora. Music sales are also highly varied across regions, some of which is also predictable.

37. Because music sales are so variable, both across tracks/albums and across regions, it is important for analyses of these sales to take advantage of known information to focus attention on the part that is not predictable – and in the case of experimental data, not predictable pre-experiment. Controlling for predictable sales patterns using pre-experimental data serves to remove much of the "noise" in music sales so that we can focus on the issue at hand – the effect that Pandora has on sales. Our analytic approach did exactly this.

38. Specifically, to estimate the effect of Pandora on music sales, we used a statistical approach that accounts for pre-experimental differences in population, sales of the experimented-on music, spins of that music on Pandora, and spins of all music on Pandora to provide an accurate (unbiased) estimate of the effect of Pandora while producing a small confidence interval to aid with inference. The specific statistical approach we used was targeted maximum likelihood estimation ("tMLE").[25] This is a statistical method applied to measure the causal

[25] Although tMLE is a relatively new approach (created in 2007), the approach has been used in multiple published studies in leading peer-reviewed academic journals and is increasingly used in biostatistics and

effect of experimental manipulations after accounting for the predictable variability of results based on pre-experimental data, and is known to produce more precise estimates than other approaches. As a check on the results, I also utilized an alternative approach (a so-called "Poisson" regression); as described in Technical Appendix B, the results of this analysis align with those of tMLE, but with the theoretically expected wider confidence intervals.[26] Both approaches are valid when a large number of observations have zero sales; an outcome that happens very often in the New MSEs. We use standard bootstrap methods to ensure valid inference, recognizing our experiments all use the same regions. We combined all experiments in the New MSEs, and separately all Catalog MSEs, to estimate the average promotional effect of Pandora across experiments. This estimate can reasonably have the interpretation of *Pandora's effect on aggregate sales* of the manipulated music.[27]

39. So as to avoid any concerns related to the particular statistical approach used, and so that our results would meet rigorous academic publication standards, we decided to use the tMLE approach prior to starting our experiments. Additional details about our analysis plan, including discussion of a small pilot experiment run to substantiate the engineering code prior to full experimentation, are contained in Technical Appendix B.

---

engineering. These publications, and our own prior experience with the estimator, substantiate that it produces very precise estimates of causal effects.

[26] In Technical Appendix B, I present results for Poisson regression, and discuss limitations of other "standard" approaches – linear models or log-linear models when analyzing sales data characterized by tremendous skew in sales, and with a large number of observations (experiment – regions) lacking any sales. The Poisson regression model produces similar results to the tMLE results presented here. I performed another check on the resilience of the findings: I ran the tMLE analysis using only the top 200 experiments by sales; the results were similar.

[27] The necessary assumption for this broader interpretation is that sales of a specific sound recording are not affected by manipulation of other music in the experiments.

40. We also supplemented the primary analyses by examining how Pandora's promotional impact varies with the extent of Pandora exposure; i.e., a kind of "dose-response" analysis.[28] Specifically, we used the ratio of Pandora spins to SoundScan sales as our index of Pandora exposure, calculating the average promotional effect of Pandora along a spectrum of exposure.[29] By this measure, Pandora would be expected to have greater promotional impact for experiments with higher exposure ratios.

Results

41. I present results for the causal effect of performance of sound recordings on Pandora. Positive estimates mean that music sales are greater when the music is spinning on Pandora, that is, Pandora is promotional; negative estimates mean that Pandora substitutes for music sales. For each group of experiments, we measured the average impact of Pandora on sales overall, as well as the impact on the sales of major label recordings and, separately, on the sales of independent label recordings. In addition, I calculate average promotional impact on a per spin basis, separately for major label recordings and other recordings.

42. **Table 3** presents the average estimated promotional effects of spinning on Pandora versus not spinning on Pandora, for the New MSEs and the Catalog MSEs. As shown in the table, spinning on Pandora increases music sales by +2.31% for music new to Pandora,

[28] Dose-response, as in, comparing different doses of Pandora exposure on sales.

[29] We defined ranges of the spins to sales ratio, ranging from zero (including all experiments, as with the primary results), 1, 10, 25, and as high as 200; fewer experiments met the higher thresholds. We used contemporaneous sales and spins (2 times spins in control regions), that is, data during the experimental period, because most of the music in the New MSEs had, by design, no spins pre experiment.

The thresholds are not mutually exclusive, so that all experiments that met the 200 spins per sale threshold were in all earlier effect estimates. For all qualifying experiments, we then recalculated the promotional impact of performances of sound recordings on Pandora.

and increases music sales by +2.66% for catalog music on Pandora. Both results are statistically significant.

43. **Table 4** presents the average promotional effect of Pandora on music sales for both major label recordings and independent label recordings. As shown in the table, Pandora increases music sales for new music from major labels by a statistically significant +2.82%.[30] For catalog music, Pandora has a promotional effect for major label recordings of +2.36% and a promotional impact on independent recordings of +3.85%.[31] These results show that, even for music that is broadly familiar, playing on Pandora increases sales of music.

44. **Figures 3a** and **3b** present the promotional impact of Pandora for experiments meeting minimum Pandora exposure thresholds, for all New MSEs and Catalog MSEs, and separately by major record company versus independent. These results provide additional evidence of a strong promotional impact of Pandora on music sales. For example, for the 409 New MSEs in which Pandora spins were at least 25 times the number of sales, Pandora is +5% promotional. For the 214 New MSEs in which Pandora spins were at least 150 times number of sales, Pandora is +15% promotional.[32] Increasing promotional impacts are evident for the Catalog MSEs, and for the New MSEs when separating Major from "Other". These results show that increased exposure on Pandora is associated with even greater promotional impact.

45. There are at least two reasons to look at the data in this fashion. First, I believe this result (i.e., increasing exposure) is consistent with past record label behavior with respect to

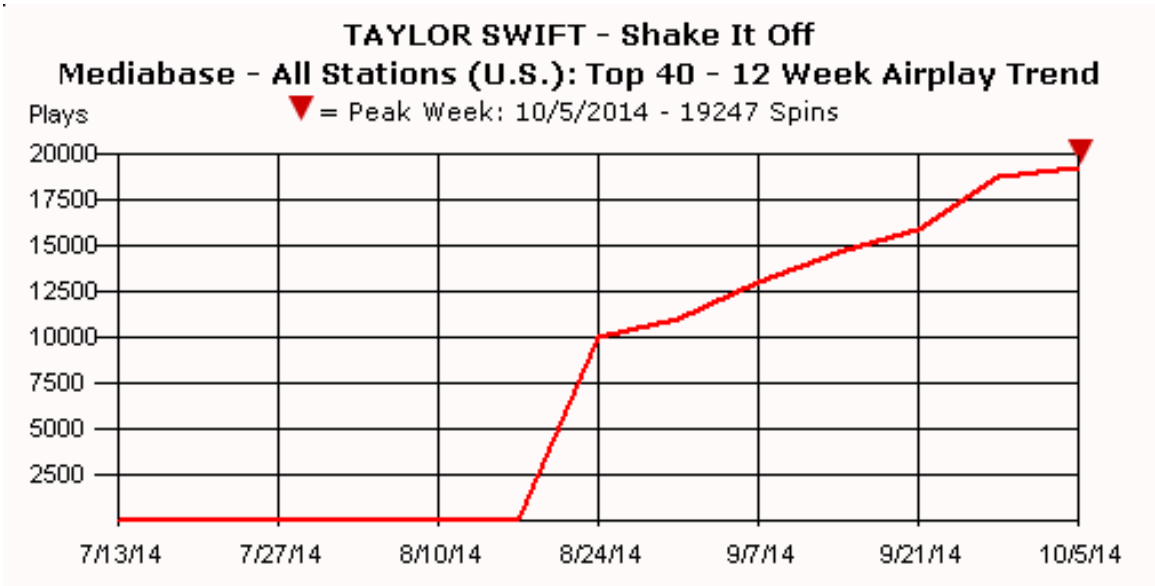
[30] For all independent label New MSEs, there is no statistically significant evidence that demonstrates whether Pandora is promotional or substitutional. There is also no evidence from the New MSEs of a difference in the promotional impact that Pandora has on the independent label recordings versus the major label recordings.

[31] The difference in promotional effect on catalog music is statistically significant (p-value: 0.002).

[32] Note that these levels of Pandora exposure are not rare: over 50% of New MSEs had at least 25 spins per sale, and over 25% of New MSEs had over 150 spins per sale.

performances on terrestrial radio. As I understand it, record labels for decades have attempted (whether through pecuniary or non-pecuniary payoffs) to influence terrestrial radio stations to play their music more frequently than the music of other labels. The record labels' attempts to increase spins relative to sales is consistent with this approach.

46. Second, I understand that it may be quite common for tracks to have significant sales before having significant spins on Pandora. For example, Pandora typically adds music to its listeners' playlists once per week on Tuesday. Because new music is often released on Tuesday, it may be that new "hits" have been playing on other platforms for a week before they are playing on Pandora. In addition, because Pandora relies on listeners to increase (or decrease) the relative spins of music, even after a track is introduced to Pandora's listeners it may take time to spin at levels equivalent to spins on other platforms, which are more able to effect immediate change in the relative spins of a particular track. For example, Taylor Swift's most recent single, "Shake It Off," was released on August 18th and immediately began playing on terrestrial radio stations across the country. In contrast, Pandora did not begin spinning "Shake It Off" until August 25th. The below graph shows that "Shake It Off" was already spinning more than 10,000 times each day on terrestrial radio stations (likely reaching an audience of millions of listeners) before it began playing on Pandora.



It is logical that Pandora would not create a detectable impact on sales for songs that are playing heavily on alternative platforms while playing very little (relatively) on Pandora.

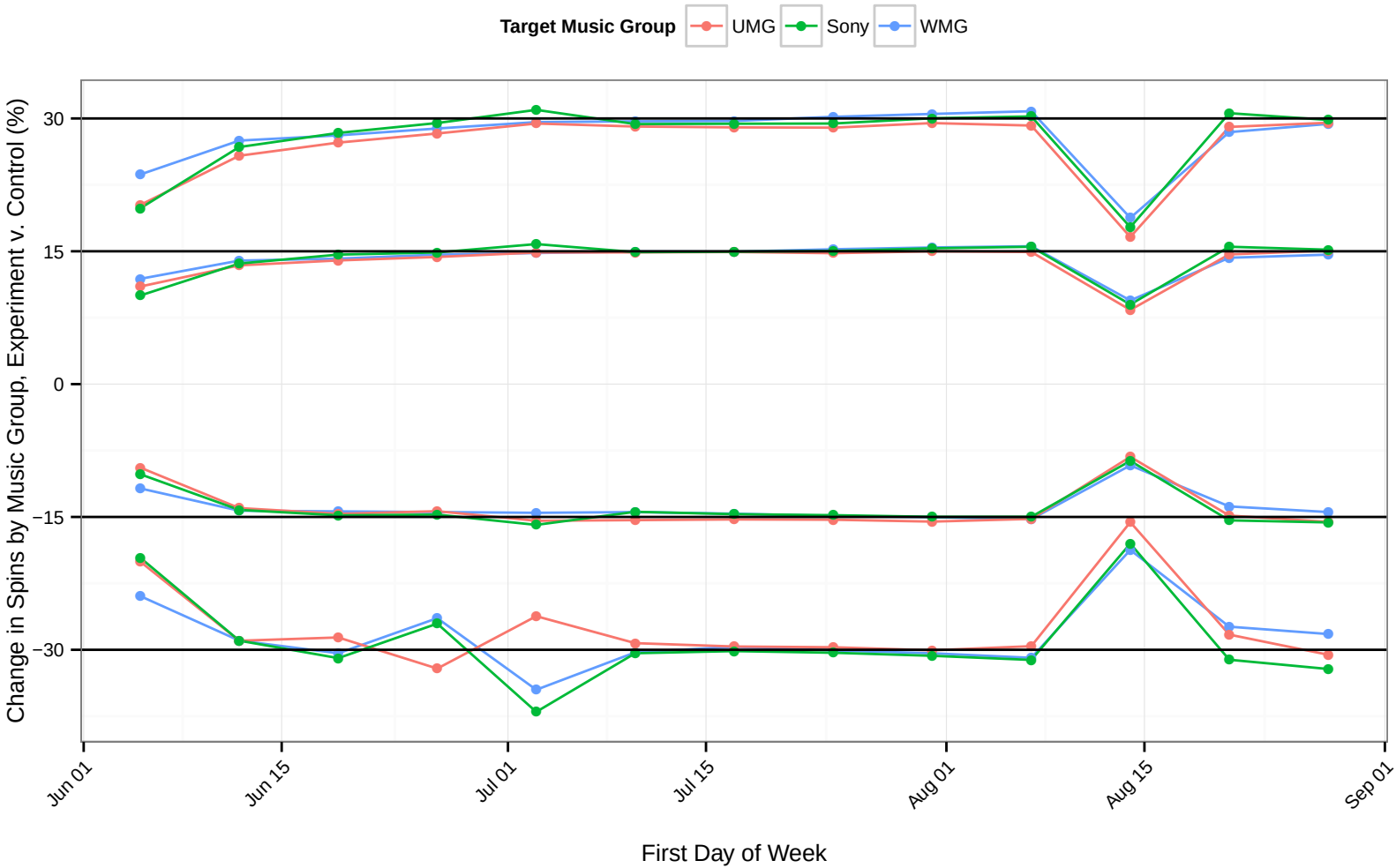
47. The above results reflect the impact of turning certain songs "on" or "off" on Pandora (*i.e.*, spinning or not spinning). Because in the normal course we spin different songs at different frequencies, turning Pandora "off" can mean a smaller or greater decrease in spins across recordings. To adjust for this fact, my colleagues and I normalized the results on a per-spin basis. **Table 5** presents the results, in revenue per spin (with revenue measured in cents). These results suggest that, on average for new music from major labels, a spin on Pandora increases music sales revenue by a statistically significant +0.16 cents. While not all of this revenue flows to content owners because of channel fees (iTune's take, for instance), this is a notable indirect transfer from Pandora to major labels and their artists. The estimated revenue on catalog music is smaller (<0.01 cents per spin), but the promotional effect is statistically significant for both major and independent label recordings.

48. **Table 6** presents results addressing any difference in promotional impact of Pandora, on a per-spin basis, between major and independent labels. From a statistical

perspective, there is no evidence of a difference in promotional impact between majors and indies per spin.

49. The Music Sales Experiments confirm that the Pandora radio service is promotional of music sales – that is, music sales are *higher* when that music plays on Pandora. As results from a well-designed gold standard experiment, they are generalizable from the specific music sampled to other new music and catalog music that spins on Pandora. We also present evidence that the promotional effect is greater for music with greater exposure on Pandora. The robustness of the study design, and the rigor applied to the analysis, distinguish these results from any previous study of the promotional impact of a radio service.

Figure 1. Evidence of Steering
Change in Spin Share vs. Control



McBride Figure 2

RESTRICTED—Subject to Protective Order in Docket No. 14-CRB-001-WR (2016-2010) (*Web IV)*

Figure 3a. Promotive Impact vs Spins per Sale
New MSEs

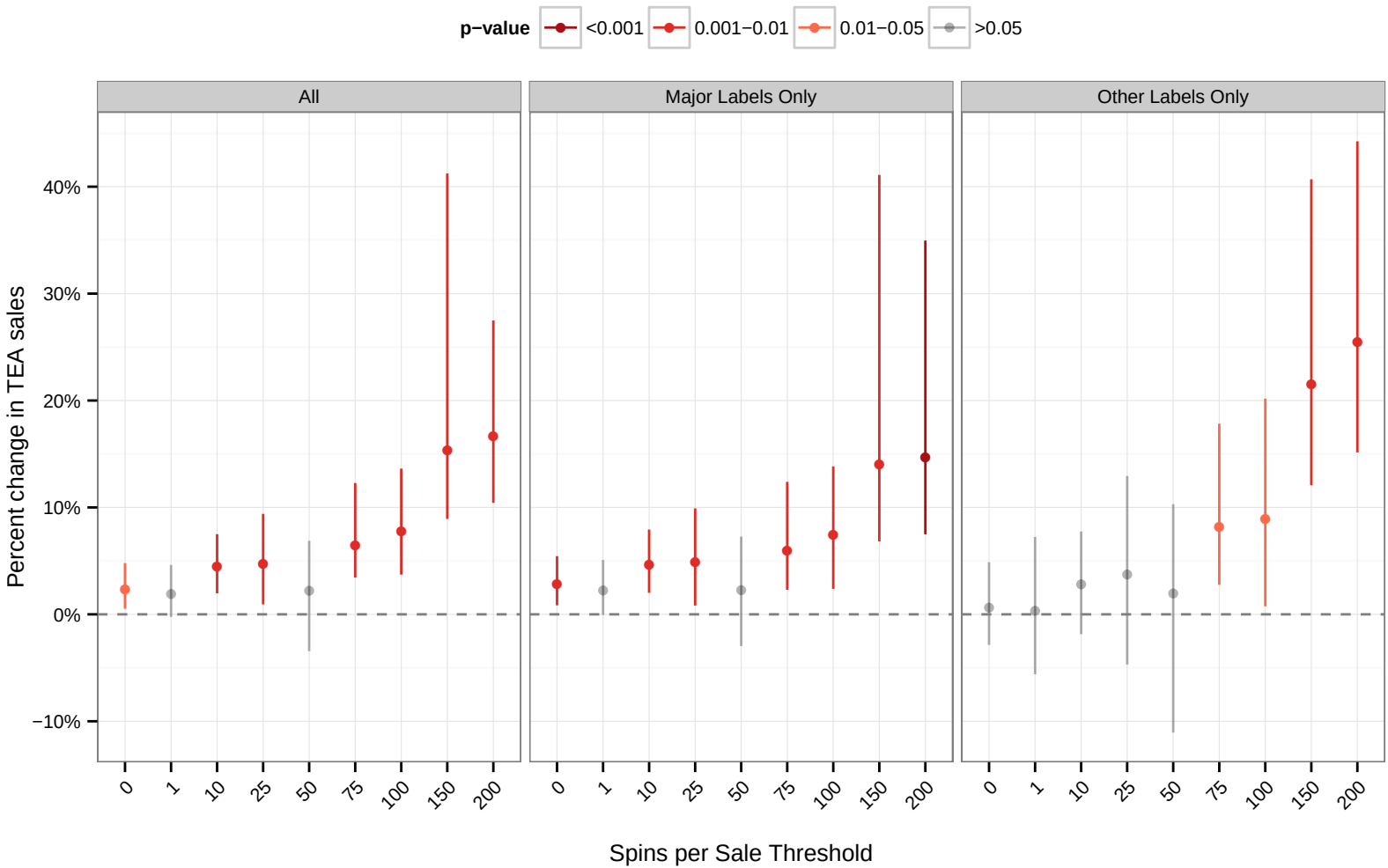


Figure 3b. Promotive Impact vs Spins per Sale
Catalog MSEs

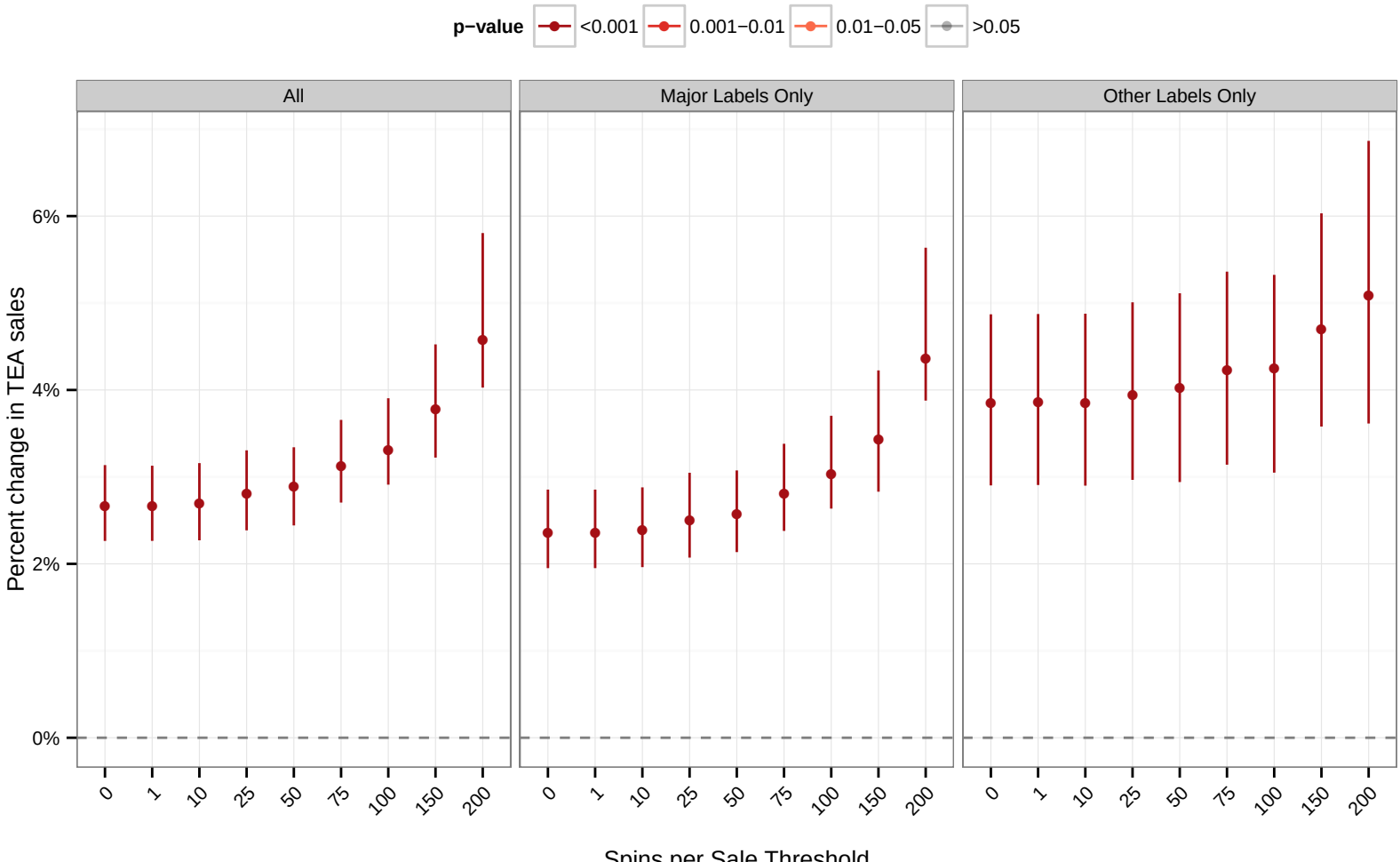


Table 1. Steering Experiments - Change in Total Hours Listened Per Registered Listener, Treatment v. Control

	-30% Steering		-15% Steering		+15% Steering		+30% Steering	
	Effect	**P-Value**	**Effect**	**P-Value**	**Effect**	**P-Value**	**Effect**	**P-Value**
Full Experimentation Period (13 Weeks)								
Target Music Group								
UMG	**-0.378%**	0.0011	-0.164%	0.1577	0.000%	0.9996	-0.124%	0.2846
Sony	-0.199%	0.0570	0.008%	0.9401	-0.030%	0.7766	**-0.260%**	0.0126
WMG	-0.119%	0.2370	0.091%	0.3652	-0.011%	0.9133	-0.022%	0.8242
Experimentation Period Excluding Week with Malfunction								
Target Music Group								
UMG	**-0.363%**	0.0017	-0.166%	0.1517	-0.011%	0.9235	-0.130%	0.2607
Sony	-0.195%	0.0610	-0.002%	0.9844	-0.028%	0.7897	**-0.258%**	0.0133
WMG	-0.111%	0.2680	0.083%	0.4069	-0.012%	0.9072	-0.019%	0.8518

Notes:

1. Results are the percent difference in hours per registered listener, treatment versus control. Negative values (that are statistically significant) indicate reduced listening. Calculated as the weighted average from weekly listening.
2. Steering Experiments ran from June 4, 2014 to September 3, 2014. Full data period for June 5, 2014 - September 3, 2014.
3. The Pandora A/B Framework malfunctioned on August 18, 19, and 20, during which no steering occurred. See Figure 1 for effect on average steering for the week
4. Effect estimates with a p-value below 0.05 are bold.

Table 2. Summary of Music Sales Experiments

	New Music Sales Experiments			Catalog Music Sales Experiments	
	Round 1	**Round 2**	**Round 3**	**Round 1**	**Round 2**
Start Date	June 10, 2014	June 24, 2014	July 8, 2014	July 16, 2014	August 12, 2014
End Date	August 5, 2014	August 19, 2014	September 3, 2014	September 9, 2014	September 9, 2014
Duration of Experiment	8 weeks	8 weeks	8 weeks	8 weeks	4 weeks
Sound Recordings Manipulated	Album	Album	Album	Song	Song
Music Selection	Music New to Pandora, Week of June 10, 2014	Music New to Pandora, Week of June 24, 2014	Music New to Pandora, Week of July 8, 2014	*Rolling Stone Top 500* Songs (Pre April 2011), random selection	*Pitchfork 500* Songs (1977-2006)
Inclusion Criteria	Sound recordings new to Pandora during selection week, plus all other tracks on the same album, and all earlier versions of the same sound recording with at least one sale in the experimentation period			≥1000 spins in June 2014 ≥25 sales previous week	≥700 spins week prior to experiment ≥25 sales previous week
Exclusion Criteria	Subsequent lead singles (a new single with the same album id as an already spinning single); greatest hits / compilations (albums with a new track that is on another, already spinning album)			Song not identified in Pandora or SoundScan data	Inclusion in Round 1 of Catalog MSEs, song not found in Pandora or SoundScan data
Number of Experiments	326	269	217	200	201
Number of Manipulated Sound Recordings	2395	1454	1082	505	451
Average Daily Spins	6750	5432	3567	25176	17568

Notes:

1. Experiment counts only include experiments with sales.
2. Average daily spins calculated as the ratio of spins during the experiment divided by the number of experiments divided by number of days in the experimentation period.

Table 3. Music Sales Experiments Results - Effect of Performance on Pandora

	New Music Sales Experiments (N=814)			Catalog Music Sales Experiments (N=401)		
	Effect	**95% CI**	**P-Value**	**Effect**	**95% CI**	**P-Value**
Effect of Performance on Pandora on Music Sales	**2.31%**	[0.52% - 4.79%]	0.012	**2.66%**	[2.26% - 3.14%]	<0.0005

Notes:

1. See Written Statement and Technical Appendix B for description of New MSEs and Catalog MSEs.
2. Estimates from targeted maximum likelihood estimation on aggregate sales during experimentation period, using pre-experimental baseline factors across experiment and region. See Technical Appendix B for additional detail.
3. Standard errors bootstrapped using 1999 replicates.

Table 4. Music Sales Experiments Results - Effect of Performance on Pandora by Music Ownership Group

	New Music Sales Experiments				Catalog Music Sales Experiments			
	Effect	**95% CI**	**P-Value**	**Number**	**Effect**	**95% CI**	**P-Value**	**Number**
Effect of Performance on Pandora on Music Sales								
Major Recording Label	**2.82%**	[0.84% - 5.44%]	0.007	124	**2.36%**	[1.95% - 2.85%]	<0.0005	294
Other Recording Label	0.62%	[-2.86% - 4.88%]	0.682	690	**3.85%**	[2.9% - 4.87%]	<0.0005	107

Notes:

1. See Written Statement and Technical Appendix B for description of New MSEs and Catalog MSEs, and identification of music ownership group.
2. Estimates from targeted maximum likelihood estimation on aggregate sales during experimentation period, using pre-experimental baseline factors across experiment and region. See Technical Appendix B for additional detail.
3. Standard errors bootstrapped using 1999 replicates.

Table 5. Music Sales Experiments Results - Revenue to Labels and Artists from Per-Spin Promotion Effect

	New Music Sales Experiments			Catalog Music Sales Experiments		
	Effect	**95% CI**	**P-Value**	**Effect**	**95% CI**	**P-Value**
Per-Spin Effect of Performance on Pandora						
Major Recording Label	**0.167 ¢ / spin**	[0.05 - 0.32]	0.008	**0.006 ¢ / spin**	[0.005 - 0.008]	<0.0005
Other Recording Label	0.048 ¢ / spin	[-0.22 - 0.36]	0.727	**0.008 ¢ / spin**	[0.006 - 0.01]	<0.0005

Notes:

1. Results are cents per spin, assuming that track-equivalent units sell for $1.
2. See Written Statement and Technical Appendix B for description of New MSEs and Catalog MSEs, and identification of music ownership group.
3. Estimates from targeted maximum likelihood estimation on aggregate sales during experimentation period, using pre-experimental baseline factors across experiment and region. *See* Technical Appendix B for additional detail.
4. Standard errors bootstrapped using 1999 replicates.

Table 6. Music Sales Experiments Results - Difference in Per-Spin Promotional Effect

	New Music Sales Experiments			Catalog Music Sales Experiments		
	Effect	**95% CI**	**P-Value**	**Effect**	**95% CI**	**P-Value**
Difference in Per-Spin Effect of Performance on Pandora						
Major vs. Other Recording Label	0.12 ¢ / spin	[-0.18 - 0.44]	0.409	-0.002 ¢ / spin	[-0.003 - 0.001]	0.113

Notes:

1. Results are cents per spin, assuming that track-equivalent units sell for $1.
2. See Written Statement and Technical Appendix B for description of New MSEs and Catalog MSEs, and identification of music ownership group.
3. Estimates from targeted maximum likelihood estimation on aggregate sales during experimentation period, using pre-experimental baseline factors across experiment and region. See Technical Appendix B for additional detail.
4. Standard errors bootstrapped using 1999 replicates.

Table 7. Music Sales Experiments Results - Effect of Performance on Pandora (Poisson Regression)

Effect of Performance on Pandora on Music Sales	New Music Sales Experiments			Catalog Music Sales Experiments		
	Effect	95% CI	P-Value	Effect	95% CI	P-Value
All Experiments	2.00%	[-0.02% - 4.49%]	0.063	**2.71%**	[1.78% - 3.71%]	<0.0005
Major Recording Label Experiments	2.55%	[-0.32% - 5.56%]	0.072	**2.37%**	[1.25% - 3.5%]	0.001
Other Recording Label Experiments	0.30%	[-3.54% - 4.29%]	0.855	**3.37%**	[0.91% - 5.13%]	0.002

Notes:

1. See Written Statement and Technical Appendix B for description of New MSEs and Catalog MSEs, and identification of music ownership group.
2. Estimates from Poisson regression pooling all New MSEs / Catalog MSEs. See Technical Appendix B for additional detail.
3. Standard errors bootstrapped using 1999 replicates.

Table 8. Music Sales Experiments Results - Effect of Performance on Pandora (Top 50 Experiments by Sales)

Effect of Performance on Pandora on Music Sales	New Music Sales Experiments				Catalog Music Sales Experiments			
	Effect	95% CI	P-Value	Number	Effect	95% CI	P-Value	Number
All Experiments	**2.22%**	[0.43% - 4.61%]	0.018	100	**3.07%**	[2.56% - 3.66%]	<0.0005	100
Major Recording Label Experiments	**2.70%**	[0.52% - 5.21%]	0.015	50	**2.77%**	[2.16% - 3.52%]	<0.0005	50
Other Recording Label Experiments	0.46%	[-3.1% - 4.8%]	0.797	50	**3.73%**	[2.81% - 4.75%]	<0.0005	50

Notes:

1. See Written Statement and Technical Appendix B for description of New MSEs and Catalog MSEs, and identification of music ownership group.
2. Estimates from targeted maximum likelihood estimation on aggregate sales during experimentation period, controlling for pre-experimental factors across experiment and region. See Technical Appendix B for additional detail.
3. Standard errors bootstrapped using 1999 replicates.

Before the
UNITED STATES COPYRIGHT ROYALTY JUDGES
THE LIBRARY OF CONGRESS
Washington, D.C.

In re	)	
DETERMINATION OF ROYALTY RATES AND TERMS FOR EPHEMERAL RECORDING AND DIGITAL PERFORMANCE OF SOUND RECORDINGS (*WEB IV*)	)	**Docket No. 14-CRB-0001-WR (2016-2020)**

DECLARATION OF STEPHAN MCBRIDE

I, Stephan McBride, declare under penalty of perjury that the statements contained in my Written Direct Testimony in the above-captioned proceeding are true and correct to the best of my knowledge, information, and belief. Executed this 7th day of October 2014 in Oakland, California.

Stephan McBride

TECHNICAL APPENDIX A: The Steering Experiments

Overview

Experimental Manipulation: the share of music played on Pandora's service based on the companies that own the sound recordings. Professor Shapiro requested specific "steering" targets: -30%, -15%, +15%, and +30%, applied to each of the experimental targets, yielding 12 total experiments: 4 steering targets x 3 experimental targets. The manipulation used was naïve – a common constant factor affecting the likelihood a song owned by the target would play, with no consideration of the listener, sound recording, or station characteristics.

Experimental Target: Music owned by one of three music ownership groups: Universal Music Group ("UMG"), Sony Music ("Sony"), or Warner Music Group ("WMG"). Sound recording ownership was determined using publicly available information acquired for a business purpose not related to any litigation or this proceeding. The ownership information is updated weekly.

Experimental Subjects: All Pandora listeners were eligible for inclusion in the experiments. Professor Shapiro directed the size of the treatment group for each of the experiments: a 5% sample of listeners for each UMG experiment, 7% sample of listeners for each Sony experiment, and 8% sample of listeners for each WMG experiment. The control group consists of a 10% random sample of listeners. All listeners are blind to their assignment.

Experimental Design: Listener-level randomization of assignment to treatment or control. The experiment was operated through the Pandora A/B Framework.

Experimental Outcome: Change in average hours listened per registered listener between each treatment group and the control group, calculated weekly and over the entire experimentation period.

Experiment Timeline: All experiments started on Wednesday, June 4, 2014; all experiments ended on Wednesday, September 3, 2014.

Study Details

1. The Steering Experiments consisted of a group of 12 experiments, each defined by a combination of a target ownership group (UMG/Sony/WMG) and a target deflection in share of spins (treatment group) as compared to spins that would occur according to the standard Pandora music recommendation results (control group). The requested spin share deflections (the "steering") were: -30%, -15%, +15%, and +30% for each of the three ownership groups manipulated.

2. The experiments all started at 3 p.m. on June 4, 2014 (a Wednesday), and all experiments ended at midnight September 4, 2014 (a Wednesday, consistent with analysis of weekly data). Consistent with Pandora practice, we tracked results using full weeks, given predictable patterns in listening across days. I excluded the data for June 4, 2014 from analysis as a partial day (3pm to 12am). Weeks ran from Thursday just after midnight to the following Wednesday at midnight, given the day the experiments started.

3. The Steering Experiments were implemented by manipulating the probability that any song from the target music group would play on Pandora. The experiment used a single manipulation factor that modified the likelihood that a sound recording owned by the target music group played, either increasing the likelihood (steering toward) or decreasing the likelihood (steering away from). The manipulation factor was originally set according to experience, and then tuned daily by comparing achieved steering versus target. The manipulation factor increased if the steering was insufficient, and decreased if the steering was

greater than target. The frequency of the manipulation factor tuning was decreased during the experiment, given noise in daily spin differences (day of week effects).[1]

Ownership Information

4. The Steering Experiments used ownership information, previously collected for Pandora business purposes. Pandora undertook a costly, lengthy, and structured program to match tracks to music groups who hold at least a partial interest in the sound recording, as Pandora had not previously collected ownership data. Pandora matched songs to music groups by matching songs to music labels and labels to music groups. Pandora used music group label lists, internal expertise and industry-recognized external sources to execute this merging process.[2] Pandora matched sound recordings to three music groups: Universal Music Group ("UMG"), Sony, and Warner Music Group ("WMG").

5. After completing the above analyses of labels and ownership information for catalog music, Pandora invested in collecting and retaining label information for each sound recording newly added to the music database. The Curation Team manually enters the ownership group (not the distributor) when ingesting new music to Pandora, using several sources to ascertain the information.[3] The ownership data are updated weekly in the Pandora

[1] The frequency of tuning was also reduced to tamp an observed pattern of alternating days of increasing then decreasing changes in the factor.

[2] Sources included: http://www.universalmusic.com/labels, http://en.wikipedia.org/wiki/List_of_Universal_Music_Group_labels, http://www.universalmusic.com/labels, http://en.wikipedia.org/wiki/List_of_Sony_Music_Entertainment_labels#Current_labels, and http://en.wikipedia.org/wiki/List_of_Warner_Music_Group_labels#Main_Labels.

[3] Sources include: Amazon.com from digital download music, and searches on CDDB / Amazon for physical CDs purchased. The Curation Team enters information from liner notes on physically ingested music if unable to identify information from searchers.

databases; each week, the steering manipulation used the then current ownership data. All results rely on the now current ownership.

Analysis

6. The Shapiro Directions stipulate that the Steering Experiments be analyzed using the standard metric for analyzing retention, namely hours per registered listener. This is the primary listener metric tracked and analyzed for internal business inquiries. The A/B Framework calculates this metric daily for 7-day periods, storing the information in databases.[4] I personally queried these databases and provided the extracted results to Professor Shapiro's team, including measures of precision (standard deviation) and number of registered listeners per experiment. The standard deviation is the standard deviation in weekly listening across all registered listeners in that experiment.

7. Professor Shapiro's team requested that I provide average hours per registered listener estimates for two periods – the entire experimental period (13 weeks), and the entire experimental period excluding the week starting August 14, 2014, due to a systematic malfunction in the A/B Framework (no steering occurred for three days). I calculated these requested outputs using the weekly data extracted from the A/B Framework, and provided the output to Professor Shapiro's team.

8. The only other data requested pertained to the spin shares for the ownership groups. For this request, I computed the spin share for the target ownership group for the 12 experiments using total spins each week. I performed no further processing of these metrics.

[4] The A/B Framework uses 7-day look back periods. The Shapiro team requested that results be presented using the first day of the week, rather than the last. I executed this change.

Results

9. Results are presented by week, consistent with Pandora's standard practice given substantial variation in listening across days, for each week in the experimental period. Statistical tests compare each treatment group to the control group; given the millions of observations, the tests use the normal distribution (based on a central limit theorem), again consistent with standard Pandora practice when analyzing results from the A/B Framework.

PUBLIC VERSION

TECHNICAL APPENDIX B: Music Sales Experiments

Overview

Experimental Manipulation: not spinning a sound recording for all Pandora listeners in randomly selected regions. For simplicity, the results are presented for the effect of spinning on Pandora, rather than the effect of not spinning on Pandora.

Experimental Target: the "New MSEs" included music new to Pandora. The manipulated music included all albums that would have started to play on Pandora, but for the experiment, along with any earlier sound recordings on the same album that had already been spinning (to include lead singles with the album). The "Catalog MSEs" included music long spinning on Pandora, randomly selected from lists of significant music (*Rolling Stone Top 500* and *Pitchfork 500*) that met minimum Pandora spin and SoundScan sale thresholds.

Experimental Subjects: All Pandora listeners with a Zip Code contained in one of the 228 SoundScan designated market areas (DMAs) with sufficient data (sales and zip code match). All listeners were blind to treatment.

Experimental Design: Region-level randomization of assignment to treatment or control applied to each experiment for all 228 SoundScan DMAs. Experimental design meets all requirements for "gold standard" design.

Experimental Outcome: Music sales during the experimentation period. We used SoundScan's own standard approach to combine album and track sales – track-equivalent albums ("TEAs"), where an album counts as 10 track sales.

Experiment Timeline: The earliest experiment started on June 10, 2014, and the final experiment ended on September 9, 2014. All experiments were eight weeks in length, except for the second round of Catalog MSEs (random sample from the *Pitchfork 500*), which were four weeks in length.

Analysis: We selected targeted maximum likelihood estimation ("tMLE") to be our statistical approach prior to experimentation. We use tMLE to estimate the average promotional effect of Pandora on sales by pooling data across the experiments. This yields an estimate that approximately weights each experiment by its total sales. This estimated effect can be interpreted as the effect of Pandora on aggregate sales of all the experiments – that is, comparing aggregate sales where all the music is available to the scenario where none of the music is available – when each experimental manipulation does not affect sales of the other manipulated sound recordings. The tMLE approach uses pre-experimental information to account for substantial variation in music sales across regions within an experiment and across experiments that is predictable. Theoretical results for this estimator establish that it is efficient – that is, has the smallest confidence interval – for the average promotional effect investigated.

Inference: We compute studentized bootstrap confidence intervals. The bootstrap simulation is based on a resampling of regions (DMAs) as the independent experimental unit. We use 1999 replicates in our presented results.

Study Details

1. The Music Sales Experiments ("MSEs") are a group of research experiments in which we prevented the performance for all listeners in randomly selected regions of either (i) recordings new to Pandora ("New MSEs"); or (ii) catalog recordings long spinning on Pandora ("Catalog MSEs"). The experimental manipulation was to *not* spin the referenced sound recordings in randomly selected regions on Pandora during the experimental period; the manipulation occurred at the region level in this experiment rather than the listener level.

2. We used SoundScan 'designed market areas' ("DMAs"), for which sales data would be available, to define the regions for experimentation. Including "sub DMAs"

(subdivisions of the five largest DMAs), SoundScan tracks sales for 230 mutually-exclusive U.S, regions. For each experiment, we randomized DMAs to the treatment group or the control group. The vast majority of Pandora listeners have provided Pandora with a zip code of residence; these zip codes were used to link listeners to DMAs based on a pre-existing Pandora database that maps zip codes to SoundScan DMAs (developed in collaboration with SoundScan previously). However, it turns out that this file combined two DMAs (Tuscaloosa and Anniston, AL into a single region). Accordingly, our randomization code jointly assigned these two spatially proximate DMAs either to the treatment or control group. We discovered this after submitting our randomization code to the core Pandora music databases; we therefore elected to remove these two DMAs from the experiments. Accordingly, we have 228 effective regions for all experiments.

3. We randomized assignment for each DMA for each experiment using a randomization algorithm external to the Pandora A/B Framework. We tested randomization by comparing the mean and variance of the randomized assignment to the theoretical results given an equal probability of assignment (as used in the actual experiments). These tests confirmed that the randomization code was operating appropriately.

4. [redacted]

[redacted]

5. [redacted]

6. [redacted]. We computed track-equivalent album ("TEA") sales, SoundScan's own measure, to combine album and single sales, counting an album sale as the equivalent of 10 track sales.

Music Inclusion

7. For each of the New MSEs rounds, we pre-selected a week for the experiments to start, and identified all albums that would newly spin on Pandora that week for potential

[redacted]

manipulation. We had no advance information about the music that would be released that week when selecting the start week.

8. The weekly approach to the New MSEs matches the weekly data available from SoundScan, and the ingestion process for new music at Pandora – music new to Pandora is made available for spinning on Tuesday each week, from sound recordings approved by Sunday at midnight. Our "all comers" inclusion rule applied to the newly approved albums (multiple tracks with the same album_id approved that week), or single tracks that did not share an album_id with any other previously spinning track. Conversely, we excluded single releases that shared an album_id with previously spinning singles, deeming follow-on singles peculiar. In particular, we were concerned that the subsequently released single – the new music – was unlikely to constitute the primary driver of music purchasing behavior during the experimentation period. As a manipulation at the album level, we decided to include all albums with lead singles that were already spinning on Pandora (in part because albums with lead singles are more popular, and we wanted to ensure our inclusion criteria captured these market entry dynamics), but not music that involved sequential singles as the most recent new music would be unlikely to comprise the maturity of listening during the experimental period.

9. In some instances, particular sound recordings on the list were not truly new to Pandora, as Pandora was already spinning a prior version of the sound recording (example: a remastered version of a catalog song). To address this situation, we performed the following steps to identify our final set of music in the New MSEs:

- Combine all new sound recordings by album
- Identify all sound recordings on the album, including those that are already spinning. Exclude the album if the newly approved music is a single, and previously spinning singles had the same album id

- If the new music is an album, include all tracks on the album in the experiment (to ensure that we integrated lead singles into the album analysis)[4]

- Check whether any specific sound recording on the album was already spinning on Pandora on a different album. If so, that album and all of its specific sound recordings were excluded from the experiment. We did this to exclude albums of Greatest Hits or partial compilation albums.[5]

10. These inclusion criteria yield a reasonable definition of music that is truly new to Pandora for the purpose of analyzing music sales at the album level. However, we went further by identifying previous versions of the songs already spinning on Pandora, and included these in the experiments as well. That is, we turned off all versions of the selected music in the randomly selected DMAs. We performed this integration to prevent the earlier track, an extremely close musicological substitute, from being played by the standard Pandora music recommendation algorithms for listeners precluded from hearing the manipulated new sound recording.

11. However, at ingestion of new music, Pandora does not record an 'album' identifier in all cases, in part because the album itself may not be available at the release of its "lead" single or singles, or there is no album – only a single is released. To distinguish these cases and to help ensure that new music that is likely from the same album was manipulated in the same regions, Pandora researchers strived to combine multiple newly approved songs by the same lead artist into one experiment even when they did not share an album id, and included any subsequent releases of new music by the same artist during the experimental period in the same

[4] By integrating all songs from the album, we include some music that was already spinning on Pandora; these songs did spin but were turned off in the experimental regions for the duration of the experiment. The entirely new to Pandora music never played on Pandora prior to the ingestion.

[5] To perform these steps, we used several data fields, including album id and song identity (an indicator of an artist-song, rather than specific sound recording). Experiments with different song identities were not excluded.

experiment. We monitored the newly approved music each Monday to match against any existing experiments; when we found music matching the inclusion rules, we included it in the earlier experiment prior to it spinning. We undertook these detailed steps to produce, as much as feasible, a clean separation of new albums for other music.

12. Thus, the New Music Sales Experiments were manipulations of the availability of all albums new to Pandora (including any lead singles), as well as previous versions of the same song by the same artist. We obtained sales data only for the new to Pandora album; any effect of the experiment on earlier albums was not assessed.

13. For the Catalog MSEs, we used third party lists of significant music from various publications to select the music for manipulation. We did this to guard against researcher selection concerns. Because these lists include familiar music, we believed that we were selecting music that, if anything, would be harder for Pandora to promote (producing smaller promotional effects). We ran two rounds of Catalog MSEs, each intended to have 200 songs, sampling first from the *Rolling Stone Top 500 Songs* (songs released before April 2011) and in the second round, selecting from the *Pitchfork 500* (based on songs released between 1977 and 2006).

14. Unlike in the New MSEs, we manipulated songs in the Catalog MSEs rather than albums, as we observed limited sales for the top 10 albums included in the *Rolling Stone Top 500* albums list,[6] a result likely attributable to increased purchases of greatest hits albums. The limited sales, and the dispersed sales across numerous albums, necessitated the change to manipulation by song rather than album. The alternative approach – including all albums with any song from the *Rolling Stone Top 500* album list – was infeasible.

[6] We only investigated sales for the top 10 albums on the *Rolling Stones Top 500 Albums* list.

15. We defined inclusion criteria for the Catalog MSEs prior to starting the experiments; we planned on randomly selecting songs for experimental manipulation from those meeting the criteria. We defined two minimum thresholds of continued use of the music – minimum spins on Pandora and minimum sales in the week preceding the experiment launch[7] – and exclusion of music in any previous MSE. [redacted] For the music in the *Rolling Stone Top 500* list, 487 songs met the inclusion criteria; for music on the *Pitchfork 500,* 201 songs met the inclusion criteria. We used a random number generator to select 200 songs for inclusion in the experiment from the eligible songs. Specifically, we used the random number generator to create a uniform random number for each eligible song; we selected the 200 songs with the highest random numbers generated in this manner. We had intended to apply the same methodology to the second round of Catalog MSEs, but when only 201 songs were eligible, we included all songs in the experiment.

16. In the Catalog MSEs, we disabled all versions of the included songs, meaning the same song and same artist (studio versions, live versions, remastered versions, etc.). [redacted]

[7] Specifically, for the first round of Catalog MSEs, a song was required to have at least 1000 spins on Pandora during June 2014, and at least 25 unit sales for the week ending July 6, 2014, to be included. For the second round of Catalog MSEs, we used a modified spin threshold (≥700 spins in the preceding week), a decision informed by the spin distribution of music on the *Pitchfork 500* list, and the shorter experimentation period (so greater Pandora exposure was deemed helpful *a priori* to detect the causal effect (≥700 per week maps to ~3000 spins per month); we applied the same ≥25 sales units in the preceding week criterion.

[redacted]

[redacted] As with the New MSEs, we monitored the newly approved music list weekly for any new versions to include in the experiment.

17. The first Catalog MSEs started on July 16, 2014 and lasted for 8 weeks. The second Catalog MSEs started on August 12, 2014 and lasted for 4 weeks. All Catalog MSEs ended on September 9, 2014.

Analysis

18. Prior to performing any analyses, we decided to estimate the causal effect of performance on Pandora by grouping experiments, rather than providing estimates for all experiments individually (>1200 effect estimates). We undertook this large number of experiments because of the predictable variability in sales for any small number of experiments; grouping of the New MSEs and, separately, the Catalog MSEs was designed to help reduce this "noise".

19. We also decided prior to analysis to separate our experiments according to Major (UMG/Sony/WMG) or "Other" (not UMG/Sony/WMG), using the same underlying ownership data used in the Steering Experiments, to assess any differential effect across ownership group.

20. The classic experimental estimator is the difference in means (sales in our case) between the treatment and control groups; it is classic because it is unbiased (on average, gives the true estimate) and easy to understand. This approach ignores all information about the predictability of sales, thus making it harder to discern the true effect of the experimental manipulation because of higher uncertainty around the estimated effect. In the present context, this imprecise estimator would not take advantage of predictable patterns, resulting in wide confidence intervals, and thus hampering the evaluation of the causal effect of performances on Pandora.

21. Prior to experimentation, we decided to use targeted maximum likelihood estimation ("tMLE"), a statistical approach that is unbiased, just like the classic estimator,[9] but one that theory establishes is more efficient (that is, produces smaller confidence intervals). For these reasons, tMLE is used to analyze experimental data (perform causal inference).[10] The tMLE achieves the efficiency gains by taking advantage of pre-experimental information in a regression framework to account for the predictable differences in sales. A second advantage of the tMLE approach is that it remains unbiased and consistent for the causal effect even if the regression model is not correctly specified.

22. There is a direct relationship between the precision of causal estimates, and how accurately we can predict sales from pre-experimental data; the more accurate are sales predictions, the more precise will be the causal effect estimate. The tMLE allows the use of powerful, automated statistical learning methods – methods that are standard in statistics - that flexibly adapt to the data to yield higher prediction accuracy than is typically achieved by fixed regression models specified, *a priori*, by a researcher. In particular, we use Lasso regression to select, in an automated way, those features that are most predictive of sales for each experiment. This is a standard algorithm, and one that removes feature determination from the researchers. This is in contrast to assuming the relevance of particular factors in predicting sales. For these reasons, prior to starting our experiments, we deemed the tMLE approach the best approach to test the research hypothesis.

[9] Further, the tMLE is unbiased for the same reason as for the classic estimator – randomized treatment assignment.

[10] Although tMLE is a new approach (created in 2007), the approach has been used in several published studies in leading peer-reviewed academic journals and it is increasingly used in biostatistics analyses of experiments, engineering, and finance. These publications, and our own prior experience with the estimator, substantiate that it produces very precise estimates of causal effects.

23. Targeted MLE builds on the classical maximum likelihood approaches, but is targeted for the parameter of interest (the average promotional effect, in our case). Traditional maximum likelihood approaches (including linear regressions) are aimed at obtaining an efficient estimate of the entire data generating distribution (in our case: sales in each region) and therefore chooses a bias-variance tradeoff appropriate for this goal. The purpose of our inquiry is not to predict sales for all 228 regions for all 1200 experiments with sales. Quite distinctly, our area of interest is to estimate the average promotional effect of Pandora across these experiments (separately for New MSEs and Catalog MSEs), which is a much lower-dimensional parameter of the entire data generating distribution than traditional methods focus on. The method is "targeted," in that it concentrates on this lower-dimensional parameter.

24. Targeted MLE consists of three steps. In the first step, we use Lasso regression to obtain predictions of sales for each experiment and each DMA. In the second step, the targeting step, we update these predictions using a Poisson regression of sales on an indicator for the treatment, using the prediction from the first step as an offset. Poisson regression is the standard model for "count data" (non-negative integer outcomes like sales),[11] and especially in the presence of a high frequency of zeros.[12] In the final step of the tMLE approach, we use the updated ("targeted") predictions for each experiment-DMA of what sales would have been in the presence or absence of Pandora – and then calculate the percent increase in predicted sales in the presence of Pandora.

[11] Technically, our data are not "count data" because of the track-equivalent album adjustment, making multiples of 10 far more frequent than would randomly arise, though they are non-negative integer data.

[12] Although classic Poisson regression calculates confidence intervals under the assumption that the (conditional) mean of sales equals the (conditional) variance of sales, this assumption is not relied on, either in this use, or later when I present Poisson regression results (because confidence intervals are estimated using bootstrap methods).

25. We directly calculate the average promotional effect for New MSEs and Catalog MSEs by pooling all experimental data. This approach is advantaged because it includes more experimental evidence than alternative approaches would, and with the assumption of no effect on sales across the experiments, the effect estimate has the interpretation of *Pandora's effect on aggregate sales* of the music. Further, this approach is comparable in design to a pooled Poisson regression (performed below as a robustness check). We considered three indirect approaches to averaging the experimental estimates: unweighted, weighted by sales, and weighted by inverse of the variance. The unweighted estimate is highly variable; the latter two approaches yield effect estimates that are close to our chosen approach. However, they have the distinct disadvantage that they cannot include experiments in which either the set of treatment regions or the set of control regions had no sales.

26. The tMLE procedure provides analytic confidence intervals. We compared these confidence intervals with those produced by bootstrap simulation and found that they are in close agreement. We performed the bootstrap simulation by treating the DMA as the independent unit of analysis, sampling a single set of 228 DMAs (with replacement) for each simulation. Consistent with treating the DMA as the unit of analysis, the data set of a given bootstrap replicate then consists of the data for all experiments corresponding to the sample DMAs. For each such bootstrap replicate we then re-computed the average promotional effect of Pandora. Given that the tMLE provides an analytic standard error for each of the bootstrap samples, we can compute *studentized* bootstrap confidence intervals using the tMLE standard error; studentized bootstrap confidence intervals have been shown to perform reliably. We elected to present the slightly larger studentized bootstrap confidence intervals rather than the analytic ones to be conservative.

27. We perform our estimation using software packages for the *R* programming language, a standard language used in data analysis. We perform our estimation using functions in base *R*, and use several of the well-regarded additional packages.[13]

28. To establish the resilience of the findings, I also calculated results from a pooled Poisson regression excluding the feature selection (Lasso) or targeting steps. Theory suggests that the reduced prediction accuracy of this approach should result in lower precision, but similar point estimates as both Poisson regression and tMLE are unbiased and appropriate for data with frequent zero sales. **Table 7** presents results from the pooled Poisson regression without targeting. Consistent with theory, the confidence intervals are larger than for tMLE; the estimated effects are however very close, providing additional evidence of the appropriate specification used in the tMLE. The point estimates are +2.00% for New MSEs, and +2.71% for Catalog MSEs for this approach; the Catalog MSEs are highly significant. These pooled Poisson results also suggest a large +3.37% promotional effect for independent catalog music.

29. If playing music on Pandora has a causal effect on sales, one might then expect to see a dose-response relationship, according to which there is a greater effect on sales in cases where there is more substantial exposure on Pandora. In the context of the large variation in sales and spin counts across experiments, a sensible metric for Pandora exposure is the ratio of Pandora spins to sales. Playing a song 10,000 times a day on Pandora might generate a noticeable lift in sales for a small artist with low baseline sales (say, 10 units per week), while similar spins on Pandora might have a lesser lift if the baseline sales are instead 500 units per week. While we did not explicitly randomize exposure (quantity of spins) across experiments, as we used an "on" / "off" treatment, we can explore this relationship by looking at the estimated

[13] *R* packages used include: *boot* package for studentized confidence intervals, and *glmnet* package for lasso regression; Poisson regression was carried out using the *glm* function included in 'base' *R*.

promotional effect along a spectrum of different exposure levels. **Figures 3a** and **3b** present the results of this analysis. We find strong evidence of increasing promotional effect from increased Pandora exposure, for both new music and catalog music.

30. One concern would be that the inclusion of many experiments in the New MSEs with very few sales materially affects the results – a variant of the concern that there are too many zeros (which makes linear and log-linear models tremendously challenged in this context). I analyze the top 50 sales experiments for Major / Indie New MSEs and top 50 sales experiments for Major / Indie Catalog MSEs to remove the smaller sales experiments. **Table 8** presents the results of this analysis; these results again reveal a statistically significant promotional effect of Pandora for all New MSEs, major label New MSEs, and all variants of the Catalog MSEs. The effect for major label catalog experiments was +2.77%, versus +3.73% for independent label Catalog MSEs.

31. Given this difference in promotional impact on Majors and Indies for catalog music, I undertook an additional inquiry to value this difference. Even with the lower promotional effect, I anticipated an increased revenue effect for Majors given higher sales, and possibly, higher retail music prices. To adjust for this expectation, my colleagues and I performed the calculation on a per-spin basis. We estimated the incremental number of TEA unit sales, across all DMAs, that Pandora was responsible for, using tMLE as before. We then divided the incremental number of unit sales by twice the number of spins in the control DMAs (given 50% chance of spinning in the experimental design). We again used bootstrap simulations to compute confidence intervals. I also compute whether there was a difference in the promotional effect between major and independent label recordings; statistically speaking, there was no evidence of a difference.

Pre Experiment

32. We performed standard statistical analyses and ran an experimental pilot prior to commencing the experiments to ensure the sound operation of our code and to inform our sample size. Specifically, we performed three checks prior to experimentation. First, we performed (statistical) power calculations to inform sample size by calculating the variance of sales for 8 researcher-selected songs, all of which were popular on Pandora and had regular sales.[14] No experimental manipulation occurred, rather, we analyzed the variability in the sales data for these songs to predict the number of experiments necessary to find a statistically significant effect of Pandora. These calculations informed the large scale of the experiments.

33. Second, we performed two "pilot" experiments from April 23, 2014, to July 22, 2014, by turning off one popular song and an entire album of medium popularity on Pandora.[15] We ran these pilots to ensure that our code operated as intended for the experiments, but also to ensure the code did not impact any other songs or the general Pandora music recommendation algorithms. Beyond the desired effect (not spinning in treated areas), we found no additional effect. These experiments were not intended to inform the selection of particular music for the experiments or to inform the design of the analyses.[16]

34. Once the selection criteria for the New MSEs was set as "all comers" on the pre-selected week, we revised the interface for entering the experimental code into the music

[14] The included songs: *Get Lucky* (Daft Punk), *Blurred Lines* (Robin Thicke), *Wake Me Up* (Avicii), *Roar* (Katy Perry), *I'm Different* (2 Chainz), *Dirt Road Diary* (Luke Bryan), *V.S.O.P.* (K. Michelle) and *Out of My League* (Fitz and the Tantrums).

[15] The song was *This is How we Roll* (Florida Georgia Line). The album was Kylie Minogue's 13 track album "Kiss Me Once."

[16] Neither the song nor the album manipulated in the pilot meets the inclusion criteria for any experiment we ran. Although we never intended to analyze these pilots for the promotional effect of Pandora, we did so at one point to test our code for computing the tMLE. The results were similar to those observed in the MSEs, but less precise as expected.

databases. The prior code, used in the pilots, was not able to scale to hundreds and potentially thousands of experiments. The team created an interface to easily create the random assignment of each experiment to all regions, but also to generate the required code to disable the playing of the music.

35. We ran standard engineering tests to confirm that the code would not undermine the Pandora listening experience.

Results

36. The primary results are in the written statement. I reiterate here that all results are presented for the promotional impact of a performance on Pandora, that is, the estimated effects are the causal effect of *spinning* on Pandora. Positive estimates mean that music sales are greater when the music is available and spinning on Pandora.[17]

37. The Music Sales Experiments confirm that the Pandora radio service is net promotional of music sales – that is, music sales are *higher* when the music plays on Pandora. As experimental results, they are generalizable from the specific music sampled to other new music and catalog music that spins on Pandora. We also present evidence that increased exposure on Pandora further enhances the promotional effect. The robustness of the study design, and the rigor applied to the analysis, distinguish these results from any previous study of the promotional impact of a radio service.

[17] The results are identical to the results of *not* spinning on Pandora. That is, the estimated effects will be the negative effect of not spinning on Pandora. Effect sizes are necessarily identical. We present the results in the affirmative for ease of interpretation.