Efficacy and Safety of Adjunctive Oral Ziprasidone for Acute Treatment of Depression in Patients With Bipolar I Disorder: A Randomized, Double-Blind, Placebo-Controlled Trial

Gary S. Sachs, MD; Kathleen S. Ice, PhD; Phillip B. Chappell, MBA, MD; Jeffrey H. Schwartz, PhD; Oksana Gurtovaya, PhD; Douglas G. Vanderburg, MPH, MD; and Bryce Kasuba, MA, MBA

ABSTRACT

Objective: To assess efficacy and safety of adjunctive ziprasidone in subjects with bipolar depression treated with lithium, lamotrigine, or valproate.

Method: 298 adult outpatients with bipolar I disorder (*DSM-IV* criteria) were randomized to receive ziprasidone, 20–80 mg twice a day, or placebo twice a day for 6 weeks plus their preexisting mood stabilizer. The primary efficacy variable was change in Montgomery-Asberg Depression Rating Scale (MADRS) total scores from baseline to 6 weeks. The key secondary efficacy endpoint was change from baseline to week 6 in Clinical Global Impressions-Severity (CGI-S) scores. Computer-administered assessments for diagnostic confidence were included for quality control and to evaluate study performance. The study was conducted between October 2007 and December 2008.

Results: The mean ± SD daily dose of ziprasidone was 89.8 ± 29.1 mg. Least squares mean ± standard error changes from baseline to week 6 on MADRS total score for ziprasidone and placebo treatment groups were −13.2 ± 1.2 and −12.9 ± 1.1, respectively, with a 2-sided *P* value of .792. There was no significant difference on the key secondary variable (CGI-S). Adjunctive ziprasidone was well tolerated. Poor quality ratings at baseline were associated with a trend for better improvement on placebo than ziprasidone. Among 43 placebo-treated subjects with poor baseline quality ratings, 29 (67.4%) had baseline MADRS scores > 10 points higher on the computer-administered assessment than the MADRS administered by the site-based rater. The response favoring placebo over ziprasidone observed in this subgroup suggests that poor signal detection in some clinical trials can be a consequence of "subject inflation" as well as "rater inflation."

Conclusions: Adjunctive ziprasidone treatment failed to separate from mood stabilizer alone on primary and secondary endpoints. Possible contributions to this result include enrollment of a substantial number of subjects with low diagnostic confidence, low quality ratings on the MADRS, and overzealous reporting of symptoms by subjects.

Trial Registration: clinicaltrials.gov Identifier: NCT00483548

J Clin Psychiatry 2011;72(10):1413–1422

Submitted: *December 22, 2009; accepted August 19, 2010.*
Online ahead of print: *May 3, 2011 (doi:10.4088/JCP.09m05934).*
Corresponding author: *Gary S. Sachs, MD, Bipolar Clinic and Research Program, Massachusetts General Hospital, 50 Staniford St, 5th Floor, Boston MA 02114 (SachsG@aol.com).*

Bipolar I disorder is a common complex, chronic illness that is associated with considerable functional impairment.[1] This dynamic, pleomorphic disorder challenges researchers as well as clinicians and, as a consequence, relatively little high quality data are available to guide clinical practice. The management of depression in patients with bipolar I disorder remains an area of significant unmet need.[2] In the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD), subjects with bipolar disorder experienced high rates of depressive relapse despite maintenance treatment with lithium, valproate, or other US Food and Drug Administration (FDA)–approved antimanic agents.[3] In view of the unmet need for adjunctive treatments for patients suffering from bipolar depression despite prescribed maintenance treatment at dosages considered adequate, we undertook a study of adjunctive ziprasidone.

Like other agents classified as atypical antipsychotics, ziprasidone is a dopamine D_2 and 5-HT_{2A} antagonist and interacts with numerous other receptors. Ziprasidone shows agonist activity at 5-HT_{1A} receptors and antagonist activity at 5-HT_{1B} and 5-HT_{1D} receptors. The affinity of ziprasidone for 5-HT_{1D} receptors, and its serotonin-norepinephrine reuptake inhibition, is comparable to that of the tricyclic antidepressant imipramine, and provides a rationale for studying ziprasidone as an antidepressant.[4] Data from prior small, open studies suggest that ziprasidone may reduce depressive symptoms associated with bipolar I disorder.[5–7]

Only 2 treatments have FDA approval for treatment of bipolar depression: the atypical antipsychotics quetiapine[8] and olanzapine-fluoxetine combination[9] have demonstrated more efficacy than placebo in reducing depressive symptoms in patients with bipolar I disorder. However, both drugs are associated with undesirable metabolic effects such as weight gain and disturbances of glucose homeostasis.[10,11] Ziprasidone has a lower propensity for weight gain and other metabolic disturbances than olanzapine or quetiapine.[12]

Adjunctive treatment with standard antidepressant medications is the most commonly prescribed intervention for patients with bipolar depression.[13] The STEP-BD showed no benefit, however, for adjunctive treatment with antidepressants (bupropion or paroxetine) compared to mood stabilizer plus placebo. To date, only one placebo-controlled study has succeeded in demonstrating the efficacy of any agent as an adjunct to lithium or valproate.[14] Although successful in an adjunct study[14] and commonly used for maintenance treatment for bipolar disorder, lamotrigine failed to separate from placebo in 5 of 5 bipolar depression monotherapy studies on primary outcome measure and 4 of 5 studies on key secondary outcome measures.[15] Another atypical antipsychotic, aripiprazole, studied for bipolar depression, has also produced negative or failed results.[16]

There are no double-blind data available to guide the care of depressed bipolar patients who have not responded to lithium, lamotrigine, or valproate. As preliminary clinical studies have suggested that

ziprasidone may have an antidepressant effect in subjects with bipolar disorder or with other psychiatric diagnoses, the present study was designed to investigate the efficacy and safety of ziprasidone as add-on therapy in patients with bipolar I disorder who were treated with lithium, valproate, or lamotrigine. In view of the frequency at which bipolar depression studies have failed or produced negative results, we incorporated an innovative computer-based rating management system into the study design.

METHOD

The study (clinicaltrials.gov registry: NCT00483548) was a randomized, double-blind, placebo-controlled, trial conducted at 78 centers located in Australia (4), India (6), and the United States (68). The protocol was approved by institutional review boards or independent ethics committees at each center, and the trial was conducted in accordance with the Declaration of Helsinki, the International Conference on Harmonisation Good Clinical Practice guidelines, and all appropriate local regulatory requirements.

The primary aim of the study was to investigate the efficacy and safety of ziprasidone as add-on adjunctive therapy in the treatment of depression associated with bipolar I disorder. Secondary objectives included examination of the effects of ziprasidone on global functioning and quality of life.

Subjects

Adult (≥ 18 years old) outpatients of either sex were eligible for the study if they had a primary diagnosis of bipolar I disorder, with the most recent episode depressed (296.5x), with or without rapid cycling, and without psychotic features, as defined in the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (*DSM-IV*).[17] The diagnosis was established by consensus between a certified site-based rater using the Mini-International Neuropsychiatric Interview[18] and an independent expert employed by Concordant Raters Systems in Boston, Massachusetts; Philadelphia, Pennsylvania; or San Francisco, California. The expert was a psychiatrist or psychologist with clinical experience and research experience who reviewed details of prior manic or mixed episodes collected directly from the subject by the computer and who validated the subjects' eligibility for randomization, if at least 1 episode met full *DSM-IV* criteria for mania or a mixed episode. The Bipolarity Index, a measure of diagnostic confidence,[19] was also used for cases in which it was not possible to confirm the diagnosis based on the computer assessment. In these cases, subjects were included only if sufficient additional diagnostic information was obtained from the investigator (or designee) to establish acceptable diagnostic confidence.[20] The onset of the depressive episode was required to be between 2 weeks and 6 months of screening. In addition, subjects were required to have a score of at least 20 on the 17-item Hamilton Depression Rating Scale (HDRS-17)[21] and a score of ≤ 12 on the Young Mania Rating Scale (YMRS)[22] at both screening and randomization.

Clinical Points

- The frequent failure of randomized controlled studies to detect differences between study medication and placebo is a significant obstacle to drug development.
- Although some studies include active comparators, this component alone does little to inform the field as to why randomized clinical trials often lack assay sensitivity.
- Using data from tandem assessments made by site-based raters and computer-administered assessments, this report examined the impact of protocol-specific eligibility criteria, diagnostic confidence, and rating quality on signal detection. The results suggest that variability in study quality can lead to study failure and that future clinical trials could benefit from procedures that do not rely exclusively on assessments made by a single rater.

Subjects were excluded from the study if they had any *DSM-IV-TR* Axis I or Axis II disorder that was clinically unstable or required treatment or if they showed ultrafast rapid cycling (defined as ≥8 mood episodes during the 12 months before screening). Other psychiatric exclusion criteria included a suicide attempt within the 3 months before screening or a score of at least 4 on the suicide item of the Montgomery-Asberg Depression Rating Scale (MADRS),[23] *DSM-IV-TR*–defined alcohol or psychoactive substance dependency within 6 months prior to screening or documented abuse of such substances within 3 months before screening, electroconvulsive therapy (ECT) within 3 months before screening, a history of nonresponse to ECT, treatment with any psychotropic medication other than lithium, valproate or lamotrigine within 1 week prior to screening, or depot neuroleptic treatment within the previous 6 months. In addition, subjects were excluded if they had clinically significant electrocardiogram (ECG) abnormalities, a history of QT interval prolongation or any medical condition or treatment that could produce such prolongation, or significant medical conditions, including a history of seizures, cardiovascular disease, neuroleptic malignant syndrome, or tardive dyskinesia that did not respond to treatment. Women of childbearing potential were required to use appropriate contraceptive precautions during the study.

Written informed consent was to be obtained before inclusion in the study. In the case of illiterate subjects, the subject provided an alternative indication, such as a thumbprint, and an impartial witness was required to provide signed confirmation that the informed consent procedure had been appropriate.

Study Design and Treatment

Study subjects comprised (1) subjects already on a mood stabilizer at screening and (2) subjects initiated on a mood stabilizer at screening. In both cases, mood stabilizer treatment had to remain stable, as defined by the protocol

requirements for lamotrigine dose (100–200 mg/d) or blood concentrations of lithium or valproate (0.6–1.2 mEq/L for lithium or 50–125 µg/mL for valproate), and was to be maintained for at least 4 weeks before randomization. Subjects whose mood stabilizer therapy had remained stable as per-protocol requirements for at least 4 weeks were randomized in a 1:1 ratio to receive adjunctive ziprasidone or placebo for 6 weeks.

Randomization was performed using a unique identification number for each subject and was stratified according to the type of mood stabilizer therapy (lithium, valproate, or lamotrigine). An internet-/telephone-based randomization and drug management system was used to provide the identification number and to assign either ziprasidone or matching placebo capsules to each subject throughout the trial. Blinding was to be broken only in the event of an emergency that required knowledge of the treatment for subject safety. One formal interim analysis was to be performed when approximately 60% of the planned subjects had either completed the study or discontinued prematurely. The Data Safety Monitoring Committee had the option to recommend stopping the study early for efficacy (nominal P value ≤ .0076, 2-sided) or for futility (nominal P value ≥.5099, 2-sided).

Subjects were instructed to take all study medication with food. The starting dose of ziprasidone was 40 mg in the evening on the day of randomization, followed by 40 mg twice daily on the second day (ie, 80 mg total daily dose). Thereafter, subjects were titrated twice daily with total daily doses in the range of 40–160 mg, depending on symptoms and tolerability. Compliance was assessed by pill counts, and blood levels of lithium and valproate were monitored via samples taken at screening, baseline, and week 6, or at the early termination visit to ensure the subject met the required therapeutic blood level specified in the protocol.

All other psychotropic medications were withdrawn at least 7 days or 4 half-lives (whichever was longer) before randomization. Lorazepam, or an alternative short-acting benzodiazepine, could be given at doses of up to 2 mg/d for up to 4 days per week during screening and the first 2 weeks of the double-blind treatment period to treat agitation or anxiety. Regulatory agency–approved nonbenzodiazepine medications could be used to treat sleep disturbances for up to 4 days per week until the end of the second week of double-blind treatment and for up to 2 days per week thereafter. The benzodiazepines and sleep agents were not to be given on the same day and were not to be used within 24 hours of efficacy assessments. Benztropine (≤ 6 mg/d) or an equivalent agent could be used to treat extrapyramidal symptoms. Propranolol (≤ 120 mg/d) could be used to treat akathisia.

Assessments

Efficacy assessments were made at baseline (randomization) and at weekly intervals thereafter. The primary efficacy endpoint was the change from baseline to week 6 in the MADRS total score. The key secondary efficacy endpoint was the change from baseline to week 6 in the Clinical Global Impressions-Severity scale (CGI-S)[24] score. Additional secondary efficacy endpoints included change from baseline in Hamilton Anxiety Rating Scale (HARS)[25] total score; change from baseline in YMRS total score; change from baseline in Global Assessment of Functioning (GAF) scale[17] score; change from baseline in Sheehan Disability Scale[26] total score; and change from baseline in Quality of Life Enjoyment and Satisfaction Scale (Q-LES-Q)[27] total score.

Only qualified raters who met educational and experience requirements participated in the trial. Prior to the start of the trial, rater training was conducted on-line and at an investigators' meeting for all participating centers. The MADRS data at each study visit were monitored using a remote site management system developed by Concordant Rater Systems, the vendor responsible for rater training and remote site management. Raters completed the training program and then received "provisional certification"; "full certification" was granted on raters demonstrating proficiency with concordance between site-based ratings and computer ratings within the acceptable concordance range over the first 3–6 actual subject ratings. Raters not meeting proficiency requirements were not allowed to enroll additional subjects.

Each site was provided with a laptop computer with the remote site management software (Concordant Rater Systems). The MADRS item scores as determined by the site-based ratings were entered on the laptop. In addition, (without assistance or input from the site rater), the subject completed an interactive interview on the computer, which selected a sequence of questions as necessary to map the subject's responses to the MADRS anchor points for each scale item. A computer-generated score was assigned based on the subject's input. Prior studies have demonstrated that site-based ratings and computer-administered MADRS are highly correlated.[28]

Item ratings scores on which the site-based ratings and computer scores differed by no more than 1 point were considered to be concordant. Concordant Rater Systems contacted raters by telephone to discuss the potential causes for discordant ratings, if the total score differential was ≥ 6 points or more than 2 items with a differential of ≥ 3 points. No further action was taken with raters who provided supporting information for their ratings; however, raters with unresolved discordance received remediation on use of appropriate probes and/or scoring conventions for the MADRS. In all cases, site-based raters were instructed not to change their original scores.

The same procedure was applied to the YMRS data at screening and baseline and to the HDRS-17 data at screening. Rater quality scores were categorized as better quality, lower quality, and poor quality if the absolute value of the difference between the computer- and site-based ratings was ≤ 5, >5 ≤ 10, or > 10, respectively. The poor quality ratings were designated rater inflation if the site-based ratings score was > 10 points higher than the computer score and subject inflation when the computer score was > 10 points higher than the site-based ratings score. Confidence in the lifetime

diagnosis of bipolar I disorder was assessed with the Bipolarity Index.[29] This scale quantifies the process suggested by Robins and Guze[30] for validating psychiatric diagnosis by scoring 5 illness domains (episode characteristics, age at onset, response to treatment, course of illness, and family history) on a 0–20 scale, on which higher scores are given to characteristics most associated with the Kraepelinian conception of bipolar disorder. Prior psychometric studies indicate that acceptable confidence for bipolar I disorder lifetime diagnosis corresponds to scores above 60 or having at least 3 domains scored 15 or higher.[20]

Safety and tolerability were assessed by recording of adverse events, physical examination, and measurement of vital signs, 12-lead ECG, and clinical laboratory evaluation. Extrapyramidal symptoms, akathisia, and dyskinesia were assessed by means of the Simpson-Angus Scale,[31] the Barnes Akathisia Scale,[32] and the Abnormal Involuntary Movement Scale (AIMS).[33]

Statistical Analysis

Statistical analyses were performed on the intent-to-treat (ITT) population, which consisted of all subjects who were randomized, received at least 1 dose of double-blind medication, and had at least 1 postbaseline primary efficacy assessment. In addition, the primary and key secondary efficacy endpoints were analyzed in the per-protocol population, which included all subjects in the ITT population with no major protocol violations.

The primary efficacy variable, the mean change in MADRS scores from baseline to week 6, was analyzed using a mixed model repeated measures (MMRM) analysis with fixed categorical effects of treatment, country, type of mood stabilizer, visit and treatment-by-visit interaction, and a fixed, continuous effect of baseline MADRS score; subject effect was included as a random effect. The mixed model repeated measures analysis used the restricted maximum likelihood estimation method, with a sandwich estimator of variance-covariance matrix of the fixed effects parameters. The analysis was performed using the SAS PROC MIXED procedure (SAS Institute Inc, Cary, North Carolina). An unstructured variance-covariance matrix was used in the REPEATED statement. Supplemental analyses of the primary endpoint included analysis of covariance (ANCOVA) of the change in MADRS scores from baseline to week 6, with missing data imputed using last observation carried forward (LOCF) principle; ANCOVA of change from baseline in MADRS scores at week 6 on observed cases only, the primary analysis using log-transformed total MADRS score; and a pattern mixture, mixed model repeated measures analysis of change from baseline in MADRS scores. The change in CGI-S score from baseline to week 6 was analyzed by mixed model repeated measures as described above, and supplementary analyses were performed by ANCOVA on both LOCF and observed cases data. Adjusted for the interim analysis, the P value threshold for the primary analysis was .0476. For change from baseline in total score for HARS and YMRS, ANCOVA similar to that for the primary endpoint was conducted at each postbaseline collection time point on the basis of both LOCF and observed cases. For change from baseline in scores for GAF, Sheehan Disability Scale, Q-LES-Q, Simpson-Angus Scale, Barnes Akathisia Scale, and AIMS, ANCOVA similar to that for the primary endpoint was conducted on the basis of the observed cases.

The sample size calculation was performed with EAST 4 software (Cytel Inc, Cambridge, Massachusetts) to account for a preplanned interim analysis, when approximately 60% of the planned number of subjects had either completed the study or discontinued prematurely. It was calculated that a sample size of 141 subjects per group (282 in total) would provide 85% power to detect a treatment difference in the mean change in MADRS scores from baseline to week 6 of 4.0 points, with a standard deviation of 11.0, using a 2-sided test at a significance level of .05.

Rating Quality Data Analysis

After completion of the efficacy analysis, the study sponsor sent unblinded treatment assignments to Concordant Rater Systems and matched with the rater quality data files. The files were reviewed for accuracy, and analyses were carried out using Stata version 11.0 statistical software.

The analysis plan compared key results from the efficacy analysis to those derived from the Rater Quality data set and evaluated a list of a priori competing hypotheses. These involved comparing results from prespecified subgroups defined by variables derived from computer-administered scales.

Quality ratings were defined based on the absolute value of the difference between the computer- and site-based ratings scores on the MADRS: better quality (difference ≤ 5), low quality (absolute value of difference from 5–10), or poor quality (difference > 10). Baseline MADRS ratings in which the site-based ratings score was > 10 higher than the computer scored were classified as indicating likely rater inflation. Baseline MADRS ratings in which the site-based ratings score was > 10 lower than the computer scored were classified as indicating likely subject inflation.

RESULTS

Between October 2007 and December 2008, 792 subjects were screened, of whom 298 were randomized and 294 (147 in each group) received treatment (Figure 1). Of the 294 who received treatment, 102 subjects discontinued treatment, mainly due to adverse events and protocol violations. Thus, 192 subjects (88 in the ziprasidone group, 104 in the placebo group) completed the study (Figure 1). The sample characteristics are summarized in Table 1.

Interim Analysis

The interim analysis was performed on 168 subjects (84 subjects in each treatment group; 59.6% of the planned final sample size). At the interim analysis, the least squares mean ± standard error (SE) changes from baseline to week

Figure 1. Study Profile

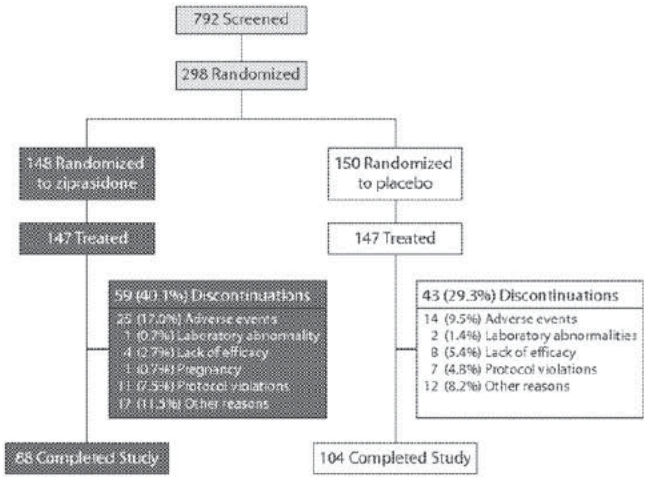

Table 1. Baseline Characteristics

Characteristic	Ziprasidone (n = 147)	Placebo (n = 147)
Sex, n (%)		
Male	58 (39.5)	56 (38.1)
Female	89 (60.5)	91 (61.9)
Race, n (%)		
White	116 (78.9)	111 (75.5)
Black	19 (12.9)	20 (13.6)
Asian	7 (4.8)	11 (7.5)
Other	5 (3.4)	5 (3.4)
Age, mean ± SD (range), y	40.4 ± 11.4 (18–64)	40.4 ± 11.9 (18–66)
Weight, mean ± SD (range), kg	84.3 ± 21.4 (45.0–156.1)	89.9 ± 23.2 (45.4–174.8)
Height, mean ± SD (range), cm	168.2 ± 10.2 (138.0–188.0)	168.0 ± 10.0 (139.7–195.0)
Time since first diagnosis of bipolar I disorder, mean (range), y	16.2 (0.07–50.7)	16.6 (0.1–45.2)
Duration of current episode, mean (range), d	76.2 (15–254)	82.9 (16–207)
No. of episodes in previous 12 mo, mean (range)	2.7 (0–20)	2.3 (0–10)
Suicidal ideation in previous 12 mo, n (%)	60 (41.4)	45 (31.5)
History of suicide attempt in previous 12 months, n (%)	6 (4.1)	4 (2.8)
Mood stabilizer, n[a]		
Lithium	53	54
Valproate	52	52
Lamotrigine	41	41

[a]One subject in the ziprasidone group did not receive a mood stabilizer during the double-blind period and was therefore excluded from the analysis.

6 on the MADRS for the ziprasidone and placebo treatment groups were −11.3 (2.18) and −13.3 (2.06), respectively, with a 2-sided *P* value of .2690 favoring placebo.

Enrollment in the study was faster than expected, and the results of the interim analysis were not available until enrollment was almost completed. On the basis of the results of the interim analysis, the Data Safety Monitoring Committee recommended that, due to study futility, already randomized subjects could complete the study but that no further subjects should enter the trial. Enrollment was completed before this recommendation was implemented.

Efficacy

Table 2 describes changes in efficacy rating scores. At baseline, there were no significant differences between the groups on any efficacy measure. The mean ± SD daily dose of ziprasidone was 89.8 ± 29.1 mg.

There was no significant difference on the primary outcome variable, the key secondary variable (CGI-S), or most of the other secondary measures, including YMRS, HARS, and Q-LES-Q. There was, however, a significant difference favoring ziprasidone over placebo on the GAF scale and the Sheehan Disability Scale.

The least squares mean ± SE change from baseline at week 6 in MADRS total score for ziprasidone- and placebo-treated subjects was −13.2 ± 1.2 and −12.9 ± 1.1, respectively (Table 2), corresponding to a least squares mean ± SE treatment difference of −0.36 ± 1.37 (95% CI, −3.07 to 2.34) that was not statistically significant (*P* = .7921). The results of the per-protocol analysis (*P* = .3989) and the sensitivity analysis

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.