2015 PC-TAR Focus Report

eDJ Group Research Report

By:

Greg Buckles, Founding Consultant

About the eDJ Group Inc.

The eDJ Group Inc. offers expert perspective, unbiased information, and pragmatic advice on eDiscovery and Information Governance products, companies, technologies and best practices. eDJ Group consultants have over 20 years professional experience in forensics, litigation, corporate governance, software design and many more related fields.

About the Author

Greg Buckles is an independent eDiscovery consultant specializing in enterprise technology and work flow solutions with over 25 years in discovery and consulting. His career spans law enforcement, legal service provider, corporate legal, law firm and legal software development. This deep and diverse background combines with exposure to the discovery challenges of Fortune 500 clients to provide a unique industry perspective.

Disclaimer:

EDJ Group is not a law firm. All expressed opinions and content are provided for general educational purposes only and are not specific legal advice, even if the author is a practicing attorney. Neither eDJ Group Inc. nor the information contained herein should be used as a substitute for competent legal advice from a licensed professional attorney in your state.

EDJ Group believes reasonable efforts have been made to ensure the accuracy of all eDJ Group Inc. original content. Content may include inaccuracies or typographical errors and may be changed or updated without notice. All eDJ Group original content is provided "AS IS" and while we endeavor to keep the information up to date and correct, we make no representations or warranties of any kind, express or implied, about the fitness for a particular purpose, completeness, accuracy, reliability, suitability, or availability with respect to the information, products, services, or related graphics for any specific purpose. Any reliance you place on such information is therefore strictly at your own risk.

In no event will EDJ Group or any of its contributors be liable for any direct, indirect, punitive, incidental, special, or consequential damages or damages for loss of profits, revenue, data, down time, or use, arising out of or in any way connected with the use of the document or performance of any services, whether based on contract, tort, negligence, strict liability or otherwise.

Contents

Overview and Purpose

This report updates the 2014 eDJ Group Analytic Adoption report with more specific usage survey questions and consumer interviews. The research focused on adoption and usage trends for **Predictive Coding - Technology Assisted Review** (PC-TAR) and other analytic technologies in the eDiscovery market. The 2014 research raised new questions that were not answered in recent surveys by other groups, so eDJ Group decided to revisit the topic to get granular answers on why consumers use analytics and how others can 'cross the chasm' to incorporate these new technologies into their eDiscovery lifecycle.

Research Methodology

Surveys: There were 41 respondents who answered the seven questions on the 2015 eDJ Group PC-TAR survey. The survey ran on the eDJ Group website for roughly three months and participants were restricted to validated eDJ Group site members. This increased the quality of the responses while reducing the volume. As you can see below, the composition of respondents is similar to the 2014 survey.
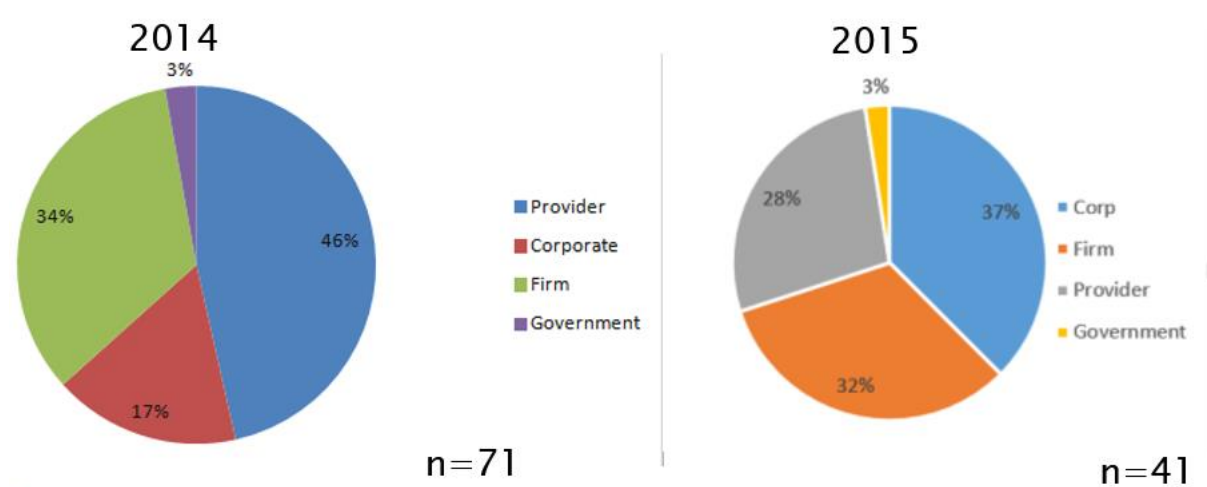

Interviews: eDJ Group conducted 15 phone and live interviews that averaged 30-60 minutes. Interviews reviewed the individual survey responses to gather more information and validate (or invalidate) the value of the question. All interview respondents had 10+ years of eDiscovery experience and had managed or conducted reviews using PC/TAR analytics. eDJ Group deliberately solicited interviews with respondents having deep expertise, market perspective and reputations as early adopters with heavy discovery/regulatory burdens.

Analytic Provider Briefings: The eDJ Group conducted briefings with analytics brand leaders and other providers with strongly differentiated usage cases or functionality to understand their evolving offerings and perceptions of the buyers. The briefings focused on critical analytic functionality, market adoption and forward looking innovation trends. eDJ Group generally conducts annual product briefings for providers with significant market share or differentiating innovation that justify the briefing time. The

objective features and subjective eDJ Notes (consultant perspectives) for all analytics providers are available to Participating Members in the eDJ Matrix[1].

2015 PC-TAR Focus

The eDiscovery market uses the term analytics for any technology that profiles, clusters, visualizes or otherwise analyzes unstructured email and file collections. Consumers buy offerings to solve a problem, not just for pretty graphs. The diagram below lays out common buying usage cases, many leveraging the same functionality for widely differing goals. The 2014 eDJ report defines basic analytic functionality and primary PC-TAR methodologies. This report attempts to shed light on what proportion of matters analytics are used in the processing and review scenarios.
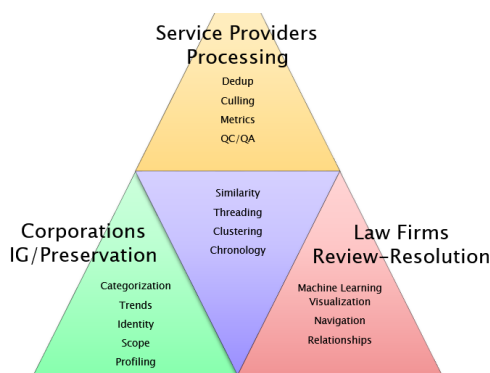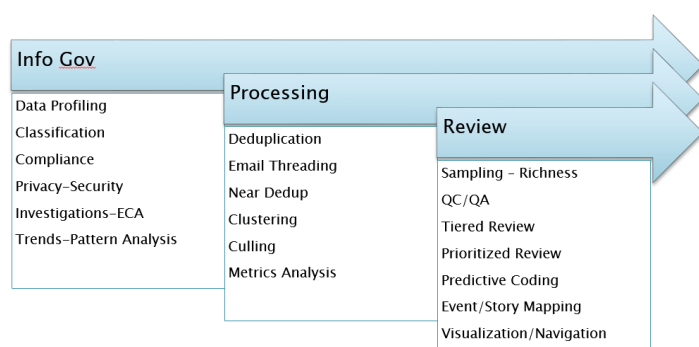
2015 Survey Results

The seven survey questions sought to understand the specific adoption rate of PC-TAR for review decisions, processing/preparation vs. review and the proportion of matters that these technologies were used in. The sixth question was a well-intentioned attempt to understand if respondents had an acceptable measurement for PC-TAR training.
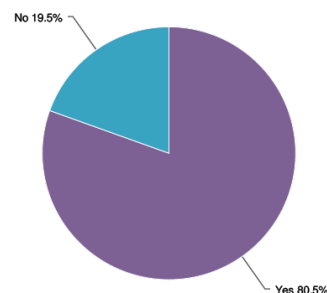
Question #1. Have you or your clients used or currently using machine learning technologies (PC-TAR) to review documents?

Even factoring in the better than average maturity of eDJ Group survey respondents, the results match up well with recent third party survey results in the 60-75% range. This was more of a control question than new ground.

Providers overwhelmingly selected volume reduction, review cost and QA, while consumers also valued strategic advantage (ECA) and discovery speed. More importantly, interviews indicated that consumers

---

[1] http://www.edjgroupinc.com/

2015 PC-TAR Focus Report

perceived review cost savings from optimized/clustered review while providers interpreted review savings as coming from PC/TAR machine learning workflows.

Question #2. What percentage of matters do you or your clients use analytics to optimize, cluster, sort or cull collections prior to linear review?

The 2014 survey did not differentiate between pre-review analytics and actual PC-TAR usage, which limited the value of the data. This year we broke the questions out (#2 & #3) to make sure that we got clear results. Respondents used a slider in 5% increments, which have been grouped into quarterly chunks for easier interpretation. Raw survey results are available online to Participating Members.
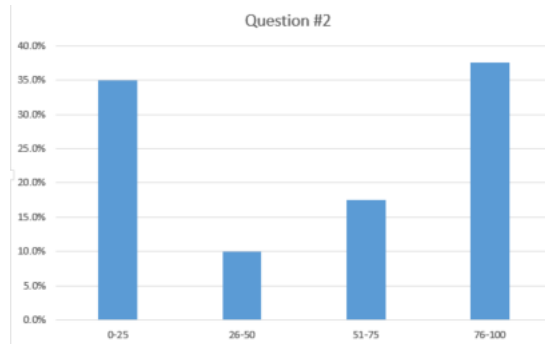
Interviews confirmed that the threading and other optimization techniques are used on matters requiring a formal, managed review process. Small reviews (<10,000 items) might skip this analytic processing when done in-house. Overall, review optimization is standard best practice and not typically disclosed or challenged by the requesting party.

Question #3. What percentage of matters do you or your clients use PC-TAR for review decisions?

Getting data on this question is the primary driver for the 2015 research. Very few respondents use PC-TAR in review for all or most of their matters. Interviews confirmed the 2014 findings that even with 'free' analytics, the effort, risk and expertise required for PC-TAR may not be justified outside of matters with extreme volumes or tight deadlines.
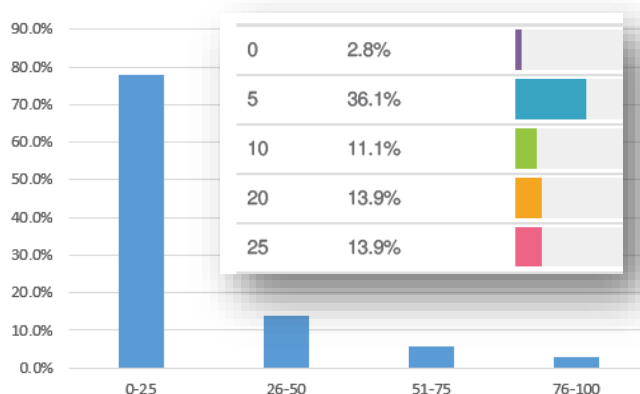
As you can see in the detailed breakdown, 50% of respondents use PC-TAR on <10% of matters. This confirms eDJ's 2014 estimate that PC-TAR was only used in 5-7% of matters.

Question #4. If you have used predictive coding, what are your plans for future use?

It is no surprise that 66% of respondents expect to increase use of PC-TAR in the future. The interviews indicated that many were more interested in using PC-TAR upstream to develop selective collection criteria rather than dramatically increasing PC-TAR for actual relevance review. The eDJ Survey numbers match up with similar recent survey results from BDO Consulting and the IGI team.
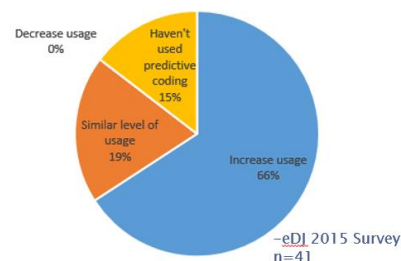
Question #5. Have you ever relied on predictive coding to make actual relevance or privilege calls on documents without counsel individually reviewing the documents?

Even though 80% of respondents have used PC-TAR for review, 44% of respondents have never relied solely on the trained relevance model to make final production decisions. That backs up interview assertions that PC-TAR is frequently used to support optimized linear review, but rarely are documents produced without human review. Interview respondents were clear that counsel was incredibly hesitant to use PC-TAR to make privilege calls, period. eDJ sampled some of the 12% of survey respondents who originally indicated that they had used PC-TAR for both relevance and privilege. When specifically questioned, each clarified that they used the system to identify pools of potential privilege items, but individual items were confirmed before being withheld for privilege.
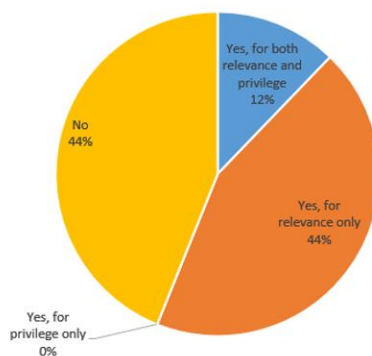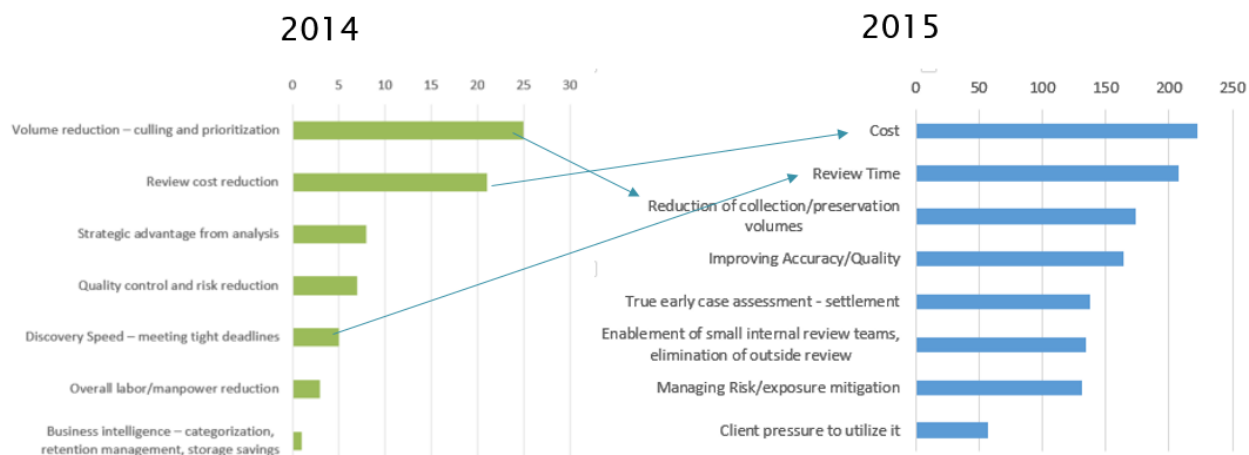
Question #6. In general, what confidence level, e.g. a 98% confidence (with a margin of error of 2) that all of the responsive document were identified, do you think is acceptable for predictive coding in civil discovery? Assume a standard deviation (margin of error) of 2.

Frankly, this survey question is based on a false or erroneous premise. Confidence level or interval is a factor used to define appropriate sample size and margin of error. The question was meant to see there were any trends or consensus among eDiscovery practitioners as to when a PC-TAR system had been trained sufficiently to be considered complete for typical relevance determination in civil discovery. Many PC-TAR workflows begin by setting this minimum confidence interval that determines the size of the training sets, but overall the question itself should be considered flawed. Interview respondents unanimously agreed that statistical measurements such as accuracy, recall, confidence interval, F1 stability and more supported counsels determination of completeness, but each was a unique decision. Respondents answered with, "It depends…"
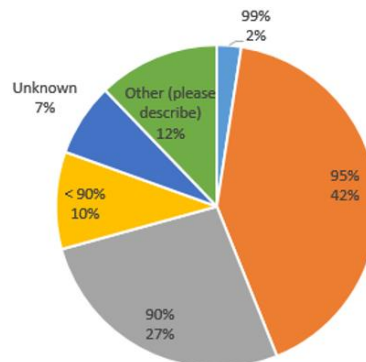
Question #7. Rank the following PC-TAR drivers in order of importance to you.

This question and follow up interviews indicated that consumer's perception of the value of PC-TAR is changing relatively quickly. Concern over ESI volume has declined in the face of new rules on proportionality and better ability to perform selective upstream collections. The 2014 report highlighted rising concern over regulatory requests with tight deadlines and heavy leverage. Interviews confirmed that short deadlines forced PC-TAR usage in some matters. Respondents did not feel that they had a methodology choice if they were going to meet the expedited requests. Rising from 5th to 2nd place as a driver for adoption in a year is a pretty drastic change.

2014

| | 0 | 5 | 10 | 15 | 20 | 25 | 30 |

Volume reduction – culling and prioritization
Review cost reduction
Strategic advantage from analysis
Quality control and risk reduction
Discovery Speed – meeting tight deadlines
Overall labor/manpower reduction
Business intelligence – categorization, retention management, storage savings

2015

| | 0 | 50 | 100 | 150 | 200 | 250 |

Cost
Review Time
Reduction of collection/preservation volumes
Improving Accuracy/Quality
True early case assessment - settlement
Enablement of small internal review teams, elimination of outside review
Managing Risk/exposure mitigation
Client pressure to utilize it

Interview Perspectives

Overall, respondents felt that PC-TAR has become another tool in their eDiscovery tools box. A tool that counsel and management frequently resist because they do not understand it's capabilities, limitations, potential costs and applications. Technology and service providers are rapidly dropping or eliminating separate PC-TAR upcharges that were making every matter a 'is it worth $X/GB to run PC-TAR?' decision. Interestingly, the strongest value propositions and success stories came from practitioners who leveraged analytics to support narrow selective collections upstream rather than uber-efficient PC-TAR review by a few trainers. Savvy, experienced practitioners have tried PC-TAR and many report that they do not think that it will replace traditional review in typical matters. The PC-TAR hype cycle has peaked and the majority of consumers have now struggled through a few high pressure PC-TAR reviews. Some expressed resentment at being sold a 'Not so easy' button. Sales reps and project managers frequently gloss over the complexities and limitations of these systems during the sales process. Others reported that they required expert support to navigate the process that frequently cancelled cost savings in what became an extended live-fire educational experience. PC-TAR systems are evolving and getting better-easier to use, but they still require knowledge that is not taught in law school, yet.

Key Interview Feedback:

- **Adoption – slowing finding right-sized usage**
  - Processing analytics, culling and prioritization are accepted without scrutiny, mature adoption with greatest value for lowest risk.
  - Most practitioners have tried PC-TAR on a few cases to address extreme volume or short deadlines. This has not led to overall adoption of PC-TAR for average matters.
  - May run PC-TAR alongside linear review for alternative perspective and QC.
  - Slowly transitioning from $/GB upcharge to all-inclusive licensing for analytics.
- **Resistance**
  - PC-TAR is seen as requiring skilled experts to do right. Leaves most practitioners hesitant to try.

- o Still seeing resistance to PC-TAR from partners and executives. Perceived as too much trouble, risk and cost unless forced to use it.
  - o "Partners are not convinced that pushing buttons leads to exhibits and good case prep."
  - o New proportionality rules may reduce collection volumes and need for PC-TAR.
  - o PC – "good at finding relevant docs, but not attractive to corporate clients as a true value"
- **Providers**
  - o Challenging to understand hidden advantages/limitations to different systems.
  - o Providers push cookie-cutter PC services that do not adapt to matters-collections.
  - o Glossing over 20-30% collection not suitable for clustering (too little/much text).
  - o "Did you spend less money? Did you get away with it?"
- **Upstream Diversions**
  - o Looking to control volumes upstream with broad preservation and smart selective collections.
  - o Stories of 70-80% collection reductions by smart scoping and tech.
- Automation and analytics will reduce demand for unskilled users, not for legal/IG professionals.
- Some regulators get/demand PC-TAR (DOJ), while others lag (FTC, SEC). Varies by region.
- Judges do a disservice by overstating capabilities and mandating without validation.
- Completeness and confidence intervals depend on requesting party and many variables. Another risk factor.

eDJ Perspective

15 years after Attenex introduced the first clustered review interface, consumers are still undecided as to whether the effort and expertise are justified in typical civil discovery. They have now tried modern PC-TAR on at least a couple cases and have threshold criteria to evaluate matters based on production deadlines, ESI volume, review resources available and the collection/production source. Many are well aware of PC-TAR issues such as unclusterable items (up to 30%), potential overproduction of non-relevant items (up to 300% in some workflows), expensive expert support, large training-sample sets (can be 15-25% of collection) and more. Marketing messaging glossed over these challenges with 'Easy Button' ads. When it is the right fit, PC-TAR delivers retrieval results that are quicker, cheaper and have better recall than traditional linear review. Consumers are wading through the 'trough of disappointment' that follows the 'Peak of Expectations' in Geoffrey Moore's technology adoption life cycle model. Providers are responding quickly to consumer demands and many now have 'all-in-one' eDiscovery processing/hosting options that include analytics at every stage without volume based upcharges. Several leading providers have rolled out educational initiatives designed to support eDiscovery and PC-TAR in traditional educational institutions. PC-TAR has been adopted, but not replaced skilled eDiscovery practitioners.