INSIGHTS · CUSTOMER AND MARKET INSIGHT

Predictive Insights in Human Resources: Generating Knowledge on Employee Retention

How organisations use data to understand and predict why employees leave, what decades of turnover research show, how predictive models are built and validated, and the ethical and practical limits of people analytics.

ModQu research article · 2,101 words · 21 references (APA 7th edition)

Employees who leave take knowledge, relationships and skills with them, and replacing them is costly. Many organisations therefore want to know who is likely to leave, why, and what might persuade them to stay. People analytics, the use of data about employees to inform decisions, promises answers. Predictive models can flag employees at risk of leaving, and analysis of survey and HR data can show which factors matter most. This article reviews what turnover research has established, how predictive insight on retention is generated, and the limits and responsibilities that come with it.

Why turnover matters

Turnover has measurable effects on organisations. Park and Shaw (2013) conducted a meta-analysis of studies relating turnover rates to organisational performance and found a negative relationship overall: organisations with higher turnover tended to perform worse, with the relationship strongest at lower levels of turnover. Hancock et al. (2013) reached a similar conclusion in another meta-analysis, finding that the negative effect of turnover on performance was stronger for some types of organisation and employee than others.

Not all turnover is harmful. The departure of poor performers, or of employees whose roles are no longer needed, may benefit the organisation. Retention insight is therefore about understanding which departures matter most, not about keeping everyone.

What turnover research shows

Research on why people leave their jobs has a long history. March and Simon (1958) proposed that the decision to leave depends on two factors: the perceived desirability of leaving, shaped largely by job satisfaction, and the perceived ease of leaving, shaped by the availability of alternatives. Mobley (1977) described a sequence of steps leading from dissatisfaction to thinking of quitting, searching for alternatives, evaluating them and finally leaving. Hom et al. (2017) reviewed a hundred years of turnover research and described how theories had evolved from these early models to include many additional influences. Holtom et al. (2008) reviewed the research of the preceding decades and noted a shift from asking why people leave to also asking why people stay, and from studying individual decisions to studying

turnover as a characteristic of units and organisations, with its own causes and consequences. Both perspectives are useful for insight work: individual-level analysis shows who is at risk, while unit-level analysis shows which teams, sites or managers have unusually high or low turnover and why.

Meta-analyses have identified the most consistent predictors. Griffeth et al. (2000) found that the strongest predictors of leaving included job satisfaction, organisational commitment, comparisons of alternatives, intentions to quit and actual job search behaviour. Rubenstein et al. (2018) conducted a larger meta-analysis covering several hundred studies and confirmed the importance of attitudes and intentions, while also finding that factors such as job embeddedness, the work environment and the employee's relationship with the organisation contributed.

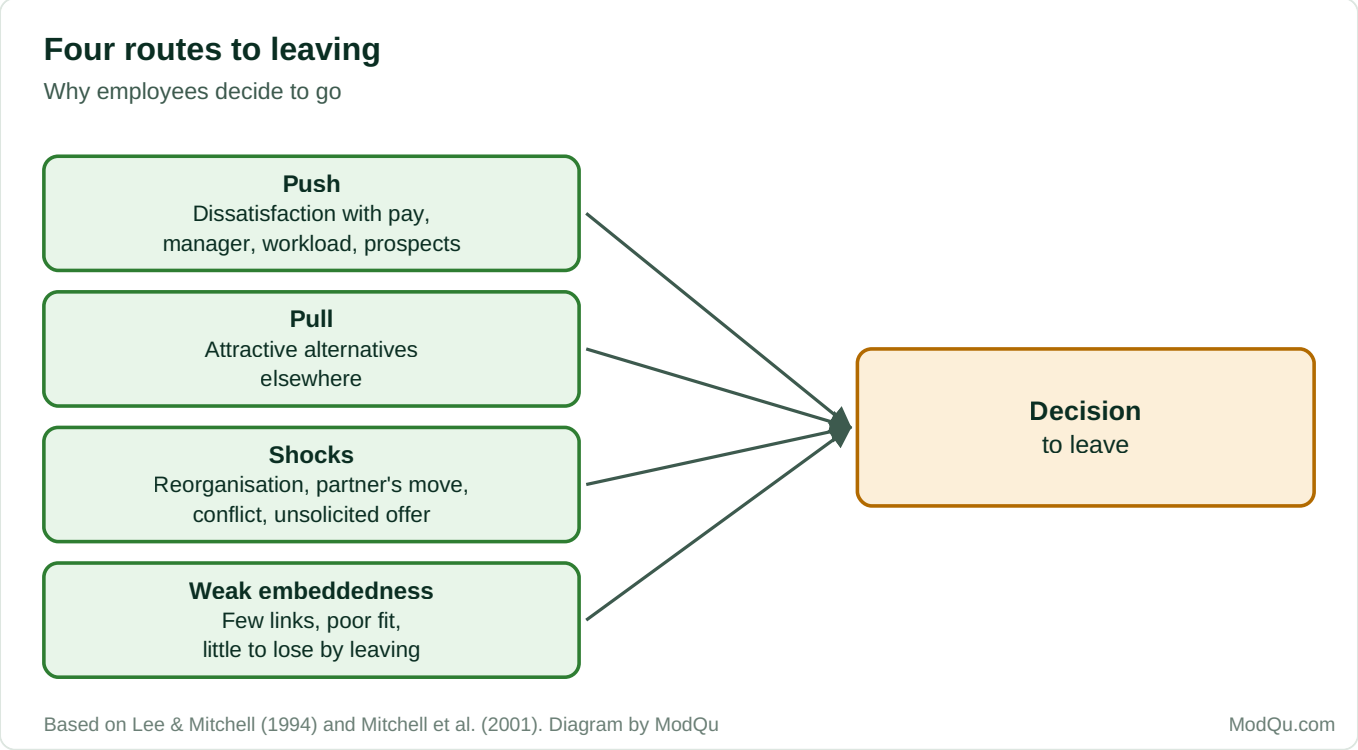


Figure 1. Four routes to leaving. Research has moved beyond dissatisfaction alone to include outside offers, sudden shocks and how strongly people are tied to their job and community. Diagram by ModQu, based on Lee and Mitchell (1994) and Mitchell et al. (2001).

Two newer ideas have been particularly influential. Lee and Mitchell (1994) proposed the "unfolding model", in which many decisions to leave are triggered by a shock, a particular event such as an unsolicited job offer, a reorganisation or a change in personal circumstances, rather than by slowly building dissatisfaction. Some people leave without searching for alternatives at all. Mitchell et al. (2001) introduced the idea of job embeddedness, the web of links, fit and potential sacrifices that keep people in their jobs, and found that it predicted leaving beyond what job satisfaction and commitment predicted. People may stay in jobs they do not love because they are tied to colleagues, community and benefits they would lose.

Timing also matters. Kammeyer-Mueller et al. (2005) followed newcomers over their first months in a job and found that changes in their attitudes and experiences over time, not only their levels at a single point, predicted whether they left. Bauer et al. (2007) conducted a meta-analysis of research on newcomer adjustment and found that role clarity, self-efficacy and social acceptance during the early period were linked to satisfaction, commitment and intentions to stay.

Hom et al. (2012) argued that research should consider not only whether employees leave but also their state of mind, distinguishing, for example, between "enthusiastic stayers", "reluctant stayers" who would like to leave but cannot, and "reluctant leavers" who are pushed out by circumstances. These distinctions matter for retention, because the actions that would keep each group differ.

Building predictive insight

People analytics combines several kinds of data: HR records such as tenure, role, pay, promotions and absence; performance data; engagement and pulse surveys; and, increasingly, data from collaboration tools and text. Marler and Boudreau (2017) reviewed research on HR analytics and found that, although interest was high, evidence on how and when it improved organisational outcomes was still limited.

A typical predictive project proceeds in several steps.

Define the outcome. Decide which departures matter: voluntary versus involuntary, regretted versus non-regretted, and within what time period.

Assemble data. Combine HR records with survey and other data, being careful to use only information that would have been available before the departure, so that the model does not "learn" from signs that the person had already decided to go.

Build and validate the model. Fit the model on past data and test it on later data that were not used in building it. Report how accurately it identifies leavers and how many false alarms it produces.

Explain the drivers. Identify which factors contribute most to predicted risk, and check them against turnover theory and against what employees themselves say.

Act and evaluate. Use the insight to change policies or to target retention efforts, and test whether these actions work.

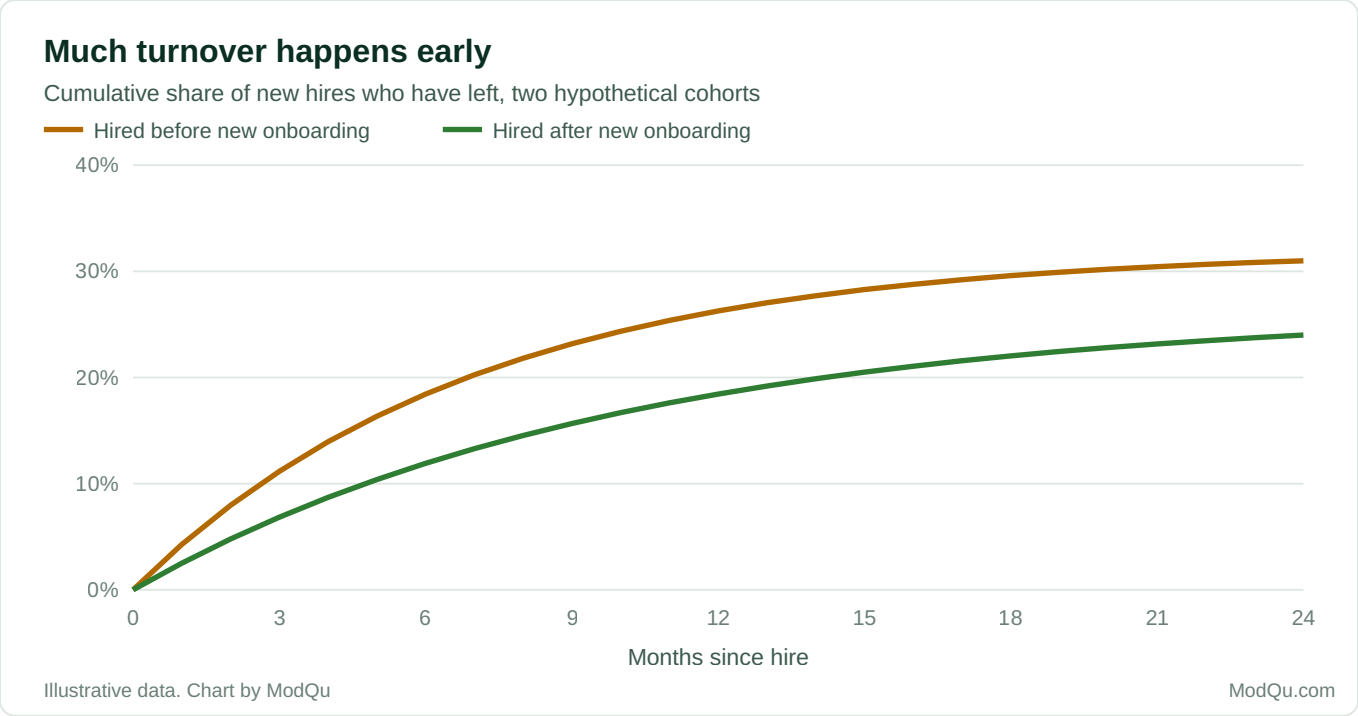


Figure 2. Cumulative leavers among two hypothetical cohorts of new hires. Much turnover happens early, which is why onboarding is a common focus of retention efforts. Illustrative data, chart by ModQu.

Prediction can also be applied earlier, at the point of hiring. Sajjadiani et al. (2019) used machine learning to analyse applicants' work histories and found that features such as the relevance of previous experience and a history of leaving jobs because of problems predicted later performance and turnover.

From prediction to retention

Knowing who is likely to leave is useful only if something can be done. Allen et al. (2010) reviewed evidence on retention and argued that many common beliefs about it are misconceptions: for example, that pay is the main reason people leave, or that turnover is mainly driven by factors outside management's control. They recommended strategies based on evidence, such as improving the quality of early experiences, developing supervisors' skills and building embeddedness.

Hausknecht et al. (2009) asked employees why they stayed and found that the reasons differed between high and low performers and between types of job: high performers more often cited advancement opportunities and organisational prestige, for example. Retention efforts are therefore most effective when they are tailored to the groups and individuals whose departure matters most, and when they address the reasons those groups give.

The same point made in research on customer churn applies here: the employees most likely to leave are not necessarily those most likely to respond to retention efforts. Testing interventions, where possible with comparison groups, shows whether they work.

Exit and stay interviews

Predictive models show patterns in data, but they rarely explain them on their own. Exit interviews and surveys, which ask departing employees why they are leaving, are the most common source of explanation. They have known weaknesses: leavers may give socially acceptable reasons, such as a better offer, rather than reasons that might burn bridges, such as a poor manager, and the reasons they give may be the final trigger rather than the underlying cause. Conducting exit interviews through a neutral party, asking about the sequence of events that led to the decision, and following up some months after departure, when people may speak more freely, all improve their value.

Stay interviews ask current employees what keeps them and what might make them leave. Because they reach people before they decide to go, they can reveal problems while there is still time to act, and they capture the reasons for staying that, as Hausknecht et al. (2009) found, differ between groups. Combining predictive models with both kinds of interview gives a fuller picture than any single source.

Common mistakes

Several mistakes recur in retention analytics. Treating all departures alike mixes regretted with non-regretted turnover and voluntary with involuntary turnover, obscuring the patterns that matter. Using data recorded after an employee had already decided to leave, such as a sudden rise in absence or a request for a reference, makes models look accurate in testing but useless in practice, because by then it is too late. Building models on small data sets, as is common in HR, produces unstable results that may not hold in the next year. And presenting risk scores to managers without explanation or guidance invites them to treat high-risk employees differently in ways that may make departure more likely.

A further mistake is to focus only on prediction. An analysis showing that employees under a particular manager leave more often is more useful than a list of individuals at risk, because it points to a cause that can be addressed. The most valuable output of retention analytics is often an insight about policies, practices or managers, not a forecast about people.

An example

Consider a hypothetical logistics company with high turnover among warehouse staff. A predictive model built on HR records shows that risk is highest in the first three months and among employees on particular shift patterns. Stay interviews with current staff and exit interviews with recent leavers explain why: new starters on rotating night shifts struggle with unpredictable schedules and receive little support from supervisors who are themselves stretched.

The company introduces fixed shift patterns for the first three months, a buddy system for new starters and training for supervisors, and tests these changes at half of its sites before rolling them out. Early turnover at the test sites falls substantially compared with the other sites. The model identified where to look; the interviews explained what was happening; and the test showed which changes worked. The company now reviews early turnover by site and shift every quarter, rather than waiting for annual figures.

Ethics and limits

Predicting individual employees' behaviour raises serious questions. Employees may not know that their data are being used to predict whether they will leave, and a high risk score could affect how they are treated, for example whether they are offered training or promotion. Tambe et al. (2019) reviewed the use of artificial intelligence in HR and identified several challenges: the complexity of HR outcomes, the small size of many HR data sets, the difficulty of measuring what matters, and concerns about fairness and the reactions of employees to algorithmic decisions.

Models can also reproduce bias. Barocas and Selbst (2016) explained how data-driven decisions can discriminate against protected groups even without intent, for example when historical data reflect past discrimination or when variables act as proxies for protected characteristics. A retention model that uses commuting distance, for instance, may indirectly reflect where people of different backgrounds live.

The expansion of workplace data collection raises further concerns. Ajunwa et al. (2017) described how employers increasingly monitor employees' activities, health and communications, often with limited legal constraints, and argued for stronger protection of worker privacy.

Several practices help. Be transparent with employees about what data are used and why. Focus predictive models on understanding patterns and improving policies, rather than on labelling individuals. Exclude variables that are protected or that could act as proxies for protected characteristics, and test models for differences in accuracy and impact between groups. Combine predictive analysis with what employees say in surveys and exit interviews. Rousseau and Barends (2011) argued that evidence-based HR practice should combine scientific research, organisational data, professional expertise and the perspectives of those affected.

Key takeaways

Employee turnover affects organisational performance, though not all turnover is harmful. Research shows that leaving is driven by dissatisfaction, attractive alternatives, sudden shocks and weak embeddedness, and that early experiences are especially important. Predictive models combining HR, performance and survey data can identify patterns and at-risk groups, but they must be validated carefully and used to improve policies rather than to label individuals. Retention efforts work best when tailored to the reasons that matter to the employees whose departure would be most costly, and when tested. Throughout, transparency, fairness and respect for employees' privacy are essential.

References

Ajunwa, I., Crawford, K., & Schultz, J. (2017). Limitless worker surveillance. *California Law Review, 105*(3), 735–776. https://doi.org/10.15779/Z38BR8MF94

Allen, D. G., Bryant, P. C., & Vardaman, J. M. (2010). Retaining talent: Replacing misconceptions with evidence-based strategies. *Academy of Management Perspectives, 24*(2), 48–64. https://doi.org/10.5465/amp.24.2.48

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671–732. https://doi.org/10.15779/Z38BG31

Bauer, T. N., Bodner, T., Erdogan, B., Truxillo, D. M., & Tucker, J. S. (2007). Newcomer adjustment during organizational socialization: A meta-analytic review of antecedents, outcomes, and methods. *Journal of Applied Psychology, 92*(3), 707–721. https://doi.org/10.1037/0021-9010.92.3.707

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management, 26*(3), 463–488. https://doi.org/10.1177/014920630002600305

Hancock, J. I., Allen, D. G., Bosco, F. A., McDaniel, K. R., & Pierce, C. A. (2013). Meta-analytic review of employee turnover as a predictor of firm performance. *Journal of Management, 39*(3), 573–603. https://doi.org/10.1177/0149206311424943

Hausknecht, J. P., Rodda, J., & Howard, M. J. (2009). Targeted employee retention: Performance-based and job-related differences in reported reasons for staying. *Human Resource Management, 48*(2), 269–288. https://doi.org/10.1002/hrm.20279

Holtom, B. C., Mitchell, T. R., Lee, T. W., & Eberly, M. B. (2008). Turnover and retention research: A glance at the past, a closer review of the present, and a venture into the future. *Academy of Management Annals, 2*(1), 231–274. https://doi.org/10.5465/19416520802211552

Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology, 102*(3), 530–545. https://doi.org/10.1037/apl0000103

Hom, P. W., Mitchell, T. R., Lee, T. W., & Griffeth, R. W. (2012). Reviewing employee turnover: Focusing on proximal withdrawal states and an expanded criterion. *Psychological Bulletin, 138*(5), 831–858. https://doi.org/10.1037/a0027983

Kammeyer-Mueller, J. D., Wanberg, C. R., Glomb, T. M., & Ahlburg, D. (2005). The role of temporal shifts in turnover processes: It's about time. *Journal of Applied Psychology, 90*(4), 644–658. https://doi.org/10.1037/0021-9010.90.4.644

Lee, T. W., & Mitchell, T. R. (1994). An alternative approach: The unfolding model of voluntary employee turnover. *Academy of Management Review, 19*(1), 51–89. https://doi.org/10.5465/amr.1994.9410122008

March, J. G., & Simon, H. A. (1958). *Organizations*. Wiley.

Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR analytics. *International Journal of Human Resource Management, 28*(1), 3–26. https://doi.org/10.1080/09585192.2016.1244699

Mitchell, T. R., Holtom, B. C., Lee, T. W., Sablynski, C. J., & Erez, M. (2001). Why people stay: Using job embeddedness to predict voluntary turnover. *Academy of Management Journal, 44*(6), 1102–1121. https://doi.org/10.5465/3069391

Mobley, W. H. (1977). Intermediate linkages in the relationship between job satisfaction and employee turnover. *Journal of Applied Psychology, 62*(2), 237–240. https://doi.org/10.1037/0021-9010.62.2.237

Park, T.-Y., & Shaw, J. D. (2013). Turnover rates and organizational performance: A meta-analysis. *Journal of Applied Psychology, 98*(2), 268–309. https://doi.org/10.1037/a0030723

Rousseau, D. M., & Barends, E. G. R. (2011). Becoming an evidence-based HR practitioner. *Human Resource Management Journal, 21*(3), 221–235. https://doi.org/10.1111/j.1748-8583.2011.00173.x

Rubenstein, A. L., Eberly, M. B., Lee, T. W., & Mitchell, T. R. (2018). Surveying the forest: A meta-analysis, moderator investigation, and future-oriented discussion of the antecedents of voluntary employee turnover. *Personnel Psychology, 71*(1), 23–65. https://doi.org/10.1111/peps.12226

Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. *Journal of Applied Psychology, 104*(10), 1207–1225. https://doi.org/10.1037/apl0000405

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review, 61*(4), 15–42. https://doi.org/10.1177/0008125619867910