EEO Studies: Statistical Methods and Data Sources

Michael P. Ward
Welch Consulting

Paper prepared for the
American Bar Association
National Conference on Equal Opportunity Law
Session: The Numbers Game: Demystifying the Use of Data in Class Actions
March 29 – April 1, 2017
Introduction

This an extremely condensed review of the many issues raised by the topic of this session. I begin by describing data sources that are typically available from employers for studying employment practices. I then briefly describe the types of statistical studies that arise in EEO litigation, the data and variables that would typically be used to fashion these studies, and the most common statistical techniques used for each. In each case I have tried to highlight some of the special issues that arise in studying each practice.

This topic is vast. If the reader has been involved in actually performing or consuming the product of any of these studies they will recognize that I have skimmed over topics that have been covered in much greater detail in many books and articles written over the last half-century. At the end I give a short bibliography of texts and some articles that I think would be a good starting point for someone new to the topic.

I. Data Sources

HRIS Transactional Data

The starting point for any study of employment practices is the “HRIS” data. (“Human Resource Information System,” a computerized data base provided by many vendors, e.g., Oracle (PeopleSoft), Kronos, SAP.) Normally, there is one system which covers the entire employee population—but not always. In complex organizations individual divisions may have their own HR system or may augment the enterprise-wide system with idiosyncratic databases designed to track items of special interest to that division. Commission structures, compensation guidelines and performance reviews may vary from one part of the company to another and not all of these are communicated to the enterprise-wide system. Alternatively, only some components are filtered upward (e.g. total review score).

These systems are structured databases (think of this as a collection of many tables/spreadsheets), which are linked to one another. The tables contain transactions—records of changes—and static information, i.e. “lookup tables” showing what different codes mean. Unfortunately, these also evolve over time.

The most important tables for EEO studies are employee histories. Employee history data is a transactional data base, or segment, containing records showing each event in an employee’s history. In most cases there is only one such history table, although there may be separate tables containing job change histories, pay histories, performance review histories, etc., which must be combined for analyses. These histories will contain the date of the event: hired, pay change, grade change, review, transfer, promotion, or termination. In order to know what actually happened on that date the analyst will compare the preceding state of the system with the change. Hire records may also show the hiring requisition that generated the hire or whether the employee was hired as part of an acquisition.

Employees may be identified by name (sometimes) and ID number. Social Security number is the preferred way of identifying people since IDs sometimes get reused and a person’s name is notoriously inaccurate. Duplicate common names, mis-spellings, nicknames, and name changes

2
make it difficult to match people over time. It is a safer bet to use the last four digits of SSN combined with name if unique IDs are not used.

Job change records are produced when any one of several employee characteristics changes:

- Status (active, on leave, type of leave)
- Job title (or job number)
- Grade (level)
- EEO category
- Location within the company (e.g. department, division, business unit)
- “reports to” (sometimes) giving the ID or name of the direct supervisor. This is a useful data item for reconstructing the hierarchy of an organization and for observing its changes over time. For supervisors and managers this item can also be used to reconstruct the number of employees reporting to a supervisor and hence a crude measure of the span of control or level of responsibility.
- Pay changes and lump sum pay records along with reason, e.g. annual merit review, promotion, bonus
- Stock awards and other forms of deferred compensation
- Leave dates and (sometimes) reasons
- Termination dates and (sometimes) reasons

Age, race, sex, name, education, marital status, dependents, address, benefit selection, and other items that change infrequently are normally kept in a separate segment of the HRIS.

More recently, companies are capturing information on prior employment and more detailed information on education—type of degree, the institution granting the degree and other post-schooling training. The spread of on-line biographic information (e.g. LinkedIn) is easing the burden of getting this information into human resource systems but, as always, this employee provided information may be inaccurate or missing entirely.

History data is preferred over “snap shots” that freeze the identity and status of employees at a given date (e.g. January 1). This is because employee turnover will cause the loss of those hired and terminated during the year. Also, the dates of pay changes, promotional progressions and performance measures will be lost. “Snap shots” can be easily created from history data if that is the chosen technique for statistical analysis.

Some warnings: HRIS data, unlike payroll data, is often a stand-alone system used principally for reporting current events—who works where now. It is not designed for litigation purposes where the questions asked are about things that happened in the sometimes distant past. These historic data may be updated and corrected as new information comes in—but not always. An early period miscode of race, gender, education level, or job may linger in HRIS since, in general, there is no need to correct historic data. The data needs to be checked extensively for anomalies like this. A common problem is that the history data records an initial hire date but only the first job in a recent “stint” of employment. Prior employment spells may or may not be retained. Some systems “backfill” information onto the initial hire record making it appear that someone has been in the same position for many years. This problem is often compounded
because people knowledgeable about earlier vintages of the data are no longer employed with the company.

Legacy HRIS systems may not be readily available. It is common for companies to switch systems and to start the new one fresh with limited data from the legacy system. This may make data in the early years of the new data unanalyzable because history data is missing. Legacy systems may be based on the same or different job/department structure.

These HRIS systems are generic. Most of them contain hundreds of potential fields many of which will be unused. Fields that are “unused” may be filled in with information completely different from the title of the field. Fields may be filled with important information but with inaccurate titles. Moreover, the use of data fields may be different in different parts of the organization. In large companies there may be separate HR management of data flowing from each division up to the enterprise.

**Miscellaneous HR Data**

- Pay ranges for jobs/grades. These may differ by geography or some other “cost of living” factor.
- Pay raise “matrices” showing recommended changes for employees falling into a range of review scores and further divided by how their salary fits into the range. For example, the highest pay change might be for someone with a low salary for their pay grade but with an outstanding review. The lowest pay change might be for someone with a low review score and a salary near the top of their range. People above the range might be eligible for a “lump sum” payment which does not affect their nominal salary.
- Succession planning documents and other assessments of readiness for promotion.

**Warnings Regarding Requests for Data**

Usually a putative class is defined in terms of location or business unit and a time frame: e.g. “Department Managers in the retail operations from January 2014 forward.” However, in order to know how a manager’s compensation came to be where it was in January of 2014 the history prior to that date is relevant (prior in-company experience and performance). Therefore, the data request should include the entire history of those who are in the putative class even though some of these records pre-date the beginning of the liability period.

**Payroll data**

These are normally organized by pay period and will show the rate of pay, amount of pay, shift differentials, hours worked, time off with and without pay, and sometimes business unit and location. Payroll data is an important adjunct to the HRIS data because it shows what people were actually paid. It is not unusual for an HRIS system to operate independently from payroll and therefore to show rates of pay that are inaccurate, time periods off work that are inaccurate, or hiring and termination dates that are inaccurate. Barring exceptional circumstances, if a check is issued to someone for work performed then the employee is employed. Likewise, if a check has not been paid the person is almost always not employed. HRIS systems may not record a termination. However, payroll is almost always determinative in gauging who and who is not
working. Exceptions include cases where salary is “continued” as part of a severance agreement even after someone has been terminated.

**Review and Performance Data**

Summary performance reviews are notoriously bland. It is not uncommon for fifty percent of employees to receive the dreaded “meets expectations” for years at a time. Pay differences and real performance differences make it apparent that this group of average performers is not all the same and, in a layoff, the weakest of these will be justifiably surprised. Some probing of the review system is usually called for. For example, many performance review systems contain both objective and subjective components which are combined into a final score: sales, customer contacts, cases resolved, subordinate employee development are often scored along with more subjective assessments. These individual components of the overall evaluation can be useful in assessing the meaning of a single summary performance measure.

Reviews often contain goals and objectives agreed to by the employee and manager during the last review cycle and may provide useful detail. The reviews may also include the employee’s own agreement or disagreement with the assessment. Warnings and improvement plans likewise help assess the overall performance of an employee. Such data are important for the analysis of pay changes and layoffs. Unfortunately, these data are often in the form of paper documents or they may be stored in the form of free form text comments in an electronic review system. The expense in capturing and quantifying this information must be weighed against its value in the ultimate statistical analysis.

**Hiring Data**

In an established firm the hiring process would begin with recruitment, either active or passive. Active recruitment might involve job fairs, visits to schools, engagement of a recruiter or outreach advertisement. Passive recruitment might involve simply posting a job opening on the company website or on a commercial or state sponsored job site. In either case, the recruitment efforts will result in “applicants,” those meeting minimum qualifications for the position. Applicant tracking systems are sometime embedded in the enterprise HRIS system, but not always. Applicants may be tested or screened for preferred qualifications and selected for one of several rounds of interview. The final selection results in an offer which may be accepted or rejected.

A variant of this process is the development of a list of potential candidates for a job or a set of jobs, who are ranked, during some stage of the selection process. This list may be maintained for several months, or even years, and candidates are drawn from the list until those viable candidates are exhausted after which a new recruitment effort is used to generate a new list. These lists may be “banded” so that everyone within a pre-determined scoring range (e.g. top 10%, next 10%, etc.) is treated the same for purposes of selection. This technique is common in public sector hiring processes. Because of the lag in the selections, by the time they are called for an interview the applicants may have moved, lost interest or be employed elsewhere. This may create a problem in determining the composition of the pool from which the selections are made.
Challenges to the hiring process can be made at any stage in the process: setting minimum qualifications, recruitment methods, testing or other screening methods, interview, and selection. The applicant tracking system will monitor flows through each stage. If protected group candidates entering the process are not selected in proportion to their representation in the initial pool it is often the defendant’s task to explain why, and where in the process, the discrepancy occurs and why this is the result of the application of non-discriminatory standards. Adverse impact may occur at any step in this selection process even though the end-to-end selection (application to hiring) may be neutral.

Challenges to minimum qualifications and testing, assuming that adverse impact has been shown, revolve around showing that these screening tools have “job validity.” Without venturing too far into the language of industrial psychology, these screening tools select for candidates with characteristics that are job related and for characteristics that are better able to predict successful candidates. Validation studies are sometimes conducted prior to the implementation of the screening tools and sometimes only after they have been challenged.

In the more typical situation, selections occur after an interview process by managers who are basing their decisions on the background and experience of the candidate as well as subjective qualifications that may only become evident in an interview. Interview notes, resumes, and application material may be helpful in reconstructing the relative ranking of candidates. Electronic applicant tracking systems for capturing stated qualifications on applications (e.g. Taleo) can ease the data collection process which might otherwise require the expensive and time-consuming processing of paper documents.

**Promotion Data**

Promotions are often a formal decision making process, but sometimes not. For the formal processes, applicant tracking systems for promotions are similar to those used for hiring. In fact, the two systems may be the same if the position being filled draws from both internal and external candidates. Typically though, an internal promotion would follow an internal notice of opening, the collection of applications, the screening of these for minimum qualification, a ranking process, interviews and selection. Applications may include data on their preferred geographic location or hours of work. These data may or may not be kept electronically.

A special problem in the analysis of promotion is the “frequent applier” problem—individuals who apply over and over to multiple positions treating applications like lottery tickets—the more tickets you buy the better your chances. In order to know if this is happening applications must be linked to the employee’s ID as opposed to simply having an application number.

In cases without a formal decision making process the first problem is determining whether a job move is a promotion. In some firms, employees may be given more responsible positions with an increase in pay but there is no indication in the HRIS system or elsewhere that this was a promotion. Here the analyst must create an empirical definition of promotion that relies on information in the HRIS system: changes in grade, job title, pay, or all three.
Layoff Data

In most cases the analysis will involve comparing the demographics of those selected to the demographics of those considered. (More on this below.) In a minority of layoff cases the layoff itself is challenged. That is, the basic economic decision of the company is up for debate. Often this occurs when a specific plant is being shut down or downsized and the argument is made that this plant was chosen because it has many older workers. Plant closure and age are not unrelated. For multi-product firms it would be common for those products that are in decline to have both low profitability and an older workforce. New and profitable products would be expanding their workforce and would tend to have a younger workforce. If workers are not substitutable across these units because of differences in skill sets, then layoffs may impact older workers disproportionately.

Workers targeted for layoff may be offered the option to transfer if their skill set matches the expanding units and the documentation of these offers will be important in explaining the overall age pattern of exits.

Layoffs involve a forward looking appraisal of employees while pre-layoff reviews are backward looking. This difference in criteria may be critical in explaining why employees with good performance reviews are let go.

In addition to basic HRIS histories, data to be collected in an analysis of layoffs would include:

- Lists of those laid off and those not laid off (e.g. WARN Act notices)
- Groupings for layoff decisions and rankings
- Pre-layoff reviews, pay raises, bonuses and other indicators of performance
- Pre-layoff training and education
- Post-layoff organization showing jobs that were retained and those eliminated
- Post-layoff hiring
- Pre-layoff attrition/retirement patterns
- Layoffs in prior waves (if any)

Pension Plan Data

Most employers now use a 401K or a similar defined contribution plan in which employer contributions are proportional to pay. However, for defined benefit plans and for other than immediate vesting of employer contributions, the plan structure can help understand retirement patterns that are important in layoff cases. For example, the typical defined benefit plan strongly incentivizes the employee to remain employed until an early retirement date but then after this date may incentivize retirement, or at least leaving that employer and working elsewhere. Managers may “save” employees from layoff if they are close to retirement age and employees themselves may volunteer for layoff if they are vested and beyond early retirement age.
Data from Paper Documents (Sampling)

Paper applications, resumes, performance reviews, hiring panel assessments, rankings (in layoffs) and resumes can be turned into analyzable data. However, because of the cost involved, sampling is normally used to create limited data for analysis. Sampling strategies are specific to the issue being addressed but here are some important observations:

1. There is no such thing as a “statistically significant sample.” The larger the sample the more statistical “power” there will be to discern the statistical significance of any factor. In general this is unknowable prior to obtaining a sample, although pre-testing with a small sample can assist in determining how large the ultimate sample must be to reach the desired level of precision.

2. Randomness is an absolute necessity. Statistical theory relies on the randomness of the sample data—or at least a method of knowing exactly the nature of the non-randomness—to invoke its mathematical results. “Taking a few from the top” is not a random sample. Neither are an “agreed 100” selected by plaintiff and defendant.

3. Random samples across the entire population are seldom statistically efficient. For example, a random sample of the entire employee population would normally result in many fewer minority employees in the sample, thereby reducing the ability of statistical tests to discern differences. It may be much more useful to sample the same number of protected group and non-protected group members to maximize the power of the statistical comparison. However, the data should be randomly selected within each group. Stratification of the sample can be both cost effective and efficient from a statistical viewpoint. For example, it may be more meaningful to sample all employees within a store from a random sample of stores.

4. Stratification may also be important for assessing “commonality” of employment practices across business units of the employer. A global random sample may well produce data that is too thinly representative of any one unit thereby impeding any statistical test designed to assess the similarity of the processes across units. Global random samples are useful only if the employment practices being investigated are assumed to be common.

5. If the document must be coded (i.e. typed in by hand) the document should be coded literally. There should be no interpretation at the point of coding. This allows for subsequent verification and testing of any interpretive re-coding of the document. For example, prior job titles on a resume should be typed in literally rather than having someone decide that it is a supervisor position or a manager position.

6. Generally, sampling is not necessary if data is available electronically. The cost of performing a statistical analysis on 100,000 employees is not 100 times as expensive the cost of analyzing 1,000. Costs rise with the complexity of the data (number of jobs, departments, businesses, changes in processes over time) but not simply with the number of people. However, large numbers of employees are often associated with large numbers of distinct jobs.
Public Data Sources

External data can be useful in establishing a general pattern in the labor market or, in a hiring case, for establishing the demographics of the external labor market. If, for example, challenges are made to an employer’s recruitment practices, then the demographics of applicants may not represent an accurate benchmark for comparison to the demographics of those hired. The classic examples are word-of-mouth recruiting or limited external advertising. External availability for hiring (at least in affirmative action planning) refers to the demographics of those possessing the requisite skill in the employer’s recruiting area. The claim is that these statistics would show what an unfettered applicant flow would look like. Estimating the demographics of this external market will require the use of Census data. Typically, these availability statistics are based on the industry or occupation of workers in an area nearby to the employer sometimes narrowed by education, age and earnings. Quite apart from their usefulness in EEO statistics these data are often the only source for establishing typical patterns of unemployment and job-finding when economic damages need to be estimated.

- Decennial U.S. Census (less useful now because of ACS)
- American Community Survey (ACS)
  - The EEO Tabulation 2006-2010 is drawn from the 5-year ACS survey
- Current Population Survey (CPS)
- Research Surveys:
  - National Longitudinal Survey of Youth,
  - Survey of Income and Program Participation (SIPP)
  - Panel Study of Income Dynamics (PDID)

II. Using the Data

Compensation Studies

Compensation studies try to explain pay differences between protected and non-protected groups. Normally this would involve hourly pay rates or base salary, but premium pay, overtime, commissions and bonuses may also be at issue.

Understanding the employer’s pay practices is a critical first step in developing a statistical model of pay. How is pay determined for a particular job title or grade? Does the employer bound pay within ranges? Are pay increases directly linked to performance reviews? How does the employer handle pay changes for demoted employees? How is starting pay determined? How do market conditions impact pay for “hot” occupations? How are exceptional performances recognized? Answers to these and many other questions will shape the list of factors included in a statistical model of pay.

There are two common statistical approaches to analyzing these pay differences. The most common are “snap shot” studies where the pay rate being studied is at a fixed point in time, e.g. January 1 salary for each of several years. For an experienced employee, salary in 2016 is the
result of decisions that reach back to the date of hire: salary at hire, and all of the raises that have been given since then. To analyze this point-in-time pay rate the factors that have influenced that pay rate should be incorporated into the statistical pay model. Thus, education, pre-hire experience, current and prior performance reviews, experience in prior positions within the employer and time in current position would be the starting point for such a study.

In the second approach, each step in the compensation process would be studied on its own: first analyze starting pay, then pay changes, bonuses, etc. This is a more focused analysis that can help locate the source of any pay discrepancy. Is it due to starting pay? Is it due to performance reviews—maybe performance reviews given in one year by one manager? The “snap shot” approach cannot answer this question easily.

Regression is almost always the statistical method used to analyze compensation differences. Regressions “control for” factors included in the model and hence net out their effect—at least in theory. Because regressions are not controlled experiments they must rely on a possibly extensive list of factors to try and net out the impact of productivity differences in arriving at a residual difference between the pay of protected and non-protected groups. Thus, models of pay for estimating, say, gender differences in pay would include variables measuring the job employees hold, time in the job, time in the company, pre-hire experience, performance in the current and past jobs, and so on. The accuracy of these methods in predicting pay can be quite weak but they are, nevertheless, interpreted as if the employees being compared are identical except for their group membership—they are not. There may be no two people in the data who are even a close match for one another. What the regression has done is to interpolate any differences between the protected and non-protected groups in their background factors and the accuracy of the resulting group pay comparison is entirely dependent on the accuracy of that interpolation.

**Promotion Studies**

If the promotion process is a formal one then the analysis of promotions would compare the representation of the protected group in the pool of qualified applicants to their representation among those selected. However, the qualifications of the candidates will be different. They may not all be applying from the same level, their experience in that lower level may be different, education, performance and other job-specific qualifications will differ. As in the analysis of pay, statistical models of promotion in these cases will attempt to control for these factors using a regression-style method in which the outcome is a selection rather than a rate of pay.

The statistical analysis will generally combine many selection decisions across many jobs. The method used should account for the fact that selections for job #1 are made from the pool of candidates for job #1, and so on through the set of jobs.

When formal promotion pools are not part of the process, statistical analysis can create an approximation to pools. A common technique is to examine the history of those who actually were promoted to a position and to see where in the company they came from: what job, location, etc. Successful candidates may be drawn from many positions. Some of these will be common “feeder jobs” for the higher level position while other jobs rarely supply successful candidates. In this case a weighted average pool may be constructed that accounts for the
representation of protected group members in each group but weights the group according to the likelihood that successful candidates come from that group. This technique has some circular reasoning embedded in it. If protected group members are disadvantaged then “feeder jobs” in which they are over represented will receive a lower weight. The resulting statistical comparison will tend to confirm parity in promotion comparisons no matter who might be favored. This can be mitigated by using protected group membership in the pools as the weighting factor.

As discussed above, the analysis of promotions should account for multiple applications by the same candidate. The simplest technique is to ask whether the person got a promotion at all, no matter how many times they applied. On the other hand, people who apply to multiple jobs may be demonstrating that they are willing to accept almost any job. Someone with narrow interest in new jobs probably will not be promoted as often as someone who is less discriminating.

As with the hiring process the multiple stages of a promotion selection process can be studied step by step. Stated minimum qualifications for a position may have adverse impact and qualifications over the minimum may not convey a real productivity advantage. Protected group members may be adversely affected at the qualifications stage, ranking, interview or selection stage of a structured promotion process and each stage can be analyzed separately.

A common misunderstanding of selection processes is to combine studies of rates of selection with studies of time to selection, as if these were two different processes. For example, it might be claimed that “not only are members of the protected group promoted at lower rates but it takes them longer to get that promotion.” In fact the two notions are intimately intertwined. They are not different things. Here’s an example. Think of the proverbial urn with 10% green balls and 90% red balls. Suppose also that a neutral process would have resulted in 50% of each. Obviously green balls are “under represented at 10%.” Now suppose that selections are made each day from the same urn. The average number of days before a red ball is selected is \(1/0.9 = 1.1\) days. The average number of days before a green ball is selected is \(1/0.1 = 10\) days. Thus green balls take longer to be selected than red balls. But this is nothing more than a mathematical consequence of the lower selection rate. It is not a distinct process with a separate adverse impact.

**Layoff Studies**

Examples might involve the downsizing of a segment of the business that has become less profitable, or an “across the board” staff reduction. Outside of a collective bargaining system where seniority governs the order of layoff, the employer is choosing to eliminate entire functions or to selectively lay off employees within a function.

Employers may establish implicit or explicit “bumping” rights in which case an eliminated function does not necessarily mean the loss of the incumbent’s job. Here the employee history data can establish whether the elimination of the function led to layoffs for all in that function.

It is seldom the case that a layoff comes without warning. Employees will know that business is slow or that a forthcoming merger will create functional redundancies. This can lead to employee behavior that will make statistical analysis of the actual layoff process complicated. In particular, it is not enough to consider only those included in the pool for layoff. Mobile
employees may leave in advance of the layoff while those who might otherwise have retired may stick around to see if severance, or other layoff benefit, is offered. We know of several instances in large layoffs where some workers volunteer to be laid off to save the jobs of younger employees.

Pre-layoff attrition, or a reduction in retirement rates, can be clues that the group of people being “considered for layoff” may not be representative of the pool of employees who might have held that job in normal times. In order to investigate these questions historic data going back prior to the layoff period may be required.

The typical statistical comparison would be made between the rate of layoff for those in the protected group (40+) and those under 40. There is on-going debate about whether this age break is the only one that should be considered and opinions differ across Courts. For example, layoff rates may be higher for those over 50 while those in the 40-50 range are laid off less often than even the under-40 group. Some of the protected group are advantaged and others may be disadvantaged. The statistical analysis may involve the computation of layoff rates in “age buckets,” for example, in 10 year age intervals.

Layoffs are presumed to be involuntary but severance plans and pension enhancements in addition to the availability of unemployment insurance benefits may induce some near their retirement age to “volunteer” for involuntary layoff. This behavior can sometimes be identified by comparing retirement rates in normal pre-layoff years to retirement rates during layoff periods.

Layoff decisions are normally made by ranking individuals within a job or grade. Those who have been in that grade the longest will tend to be the oldest while the newest in that grade are the youngest. In hierarchical organization those longest in the grade will be those passed over for advancement and the youngest will be the fast-movers. Higher performing older workers will be in higher grades. This creates a bias in the assessment of potential age discrimination since a within-grade study will almost always find age disparities disfavoring older workers. This phenomenon is sometime known as the “Peter principle” in management theory. Statistical analyses which strip away the grade-based selection criteria can help to explain this pattern but there may still be remaining age differences.

At some point obsolescence may become the principle explanation. Economic theory shows that individuals will tend to reduce their accumulation of human capital as they age since the period over which the return may be earned is shrinking, while the implicit cost of training (foregone earnings) is increasing. Thus, newer knowledge is embodied in younger workers. This is not to say that the older worker is not as productive at their skills as they have always been, but simply that the economic value of those skills may have declined. Many companies offer in-house training and tuition reimbursement subsidies. The utilization of these resources by workers of different ages can help identify those workers who are re-tooling and those who are not.

The statistical methods used in layoff studies are similar to those used in promotion studies. The relative likelihood of a layoff is a function of occupation group and measures of skill and knowledge.
Suggested Introductory Readings


