

# Washington University School of Medicine

## Gender Pay Equity Committee Final Report

April 10, 2002

### Introduction

In November 1997, the Faculty Senate Council established a university-wide Gender Pay Equity Committee to review the progress of the University toward pay equity for faculty since the last study in 1990. The committee included members from across the University. Representatives from the School of Medicine were Marion Peters, M.D.; Linda Pike, Ph.D.; and Brian Suarez, Ph.D. Jean Ensminger, Ph.D., initially chaired the committee. Upon her departure from the University, Joseph O'Sullivan, Ph.D., assumed the chairmanship of the committee.

Because of the large number of factors that contribute to the compensation levels of faculty members at the School of Medicine, the Gender Pay Equity Committee decided that a pilot study be carried out on a small sample to identify the most important variables for prediction of compensation levels. The specific purpose of the pilot study was to determine whether inclusion of factors associated with productivity, such as clinical income or grant support, would improve the fit of the model to faculty salaries.

The School of Medicine pilot study was carried out during the fall of 1999. Information on the number of publications, patents, awards etc. was obtained from the faculty in a single clinical department (the "pilot department") through self-reporting on a form provided to the faculty member. The response rate was 100%. Additional demographic data as well as financial information were obtained from the Human Resources database.

The findings of the pilot study were presented to the Gender Pay Equity Committee in December 1999. The results of the pilot study, based on the faculty in the "pilot department," indicated that no additional predictive value was achieved by adding those measures of productivity to a model that already incorporated variables such as rank, track, degree, and seniority.

Pursuant to these findings, the Gender Pay Equity Committee requested that a full pay equity study be performed at the School of Medicine using a model incorporating variables such as rank and track, which are readily available in computer databases. Accordingly, Dean William A. Peck appointed the School of Medicine Gender Pay Equity (GPE) Committee. Members of the committee are: Philip Stahl, Ph.D., chair; Barbara Cant (Human Resources); Lynn Cornelius, M.D.; Diana Gray, M.D.; Linda Pike, Ph.D.; Michael Province, Ph.D.; D.C. Rao, Ph.D.; Marilyn Siegel, M.D.; and Charles Zorumski, M.D.

The GPE Committee met in March 2000 and decided that the study should be performed on the most recent data available, which was FY2000. Since many complex, inter-correlated factors affect compensation, it was decided to take a two-stage approach to the question of gender pay equity. In the first stage, a parsimonious gender-neutral regression model was developed using data on both genders. The goal at this stage was to produce a highly parsimonious model, i.e., one which provides high predictive power ( $R^2$ ) with the fewest numbers of predictors (degrees of freedom). In the second stage, this parsimonious model was used to predict compensation for each faculty member, and the results were tested statistically for gender differences in the deviations between actual and predicted compensation. The first parsimonious model was called the "*Basic Model*." It considered basic predictors like track,

rank, degree and seniority, following a similar approach to the gender pay equity analysis done for the other schools at Washington University. The Basic Model included 22 variables (degrees of freedom, DF) and explained 77% of the variability in the compensation levels ( $R^2 = 0.766$ ). Using this model to predict compensation, there was a statistically significant gender difference in the deviations between predicted and actual compensation levels, with women being paid significantly less than men ( $P < 0.006$  overall for the medical school). Further investigation suggested that the gender differences were largely confined to women full professors on the investigator track ( $P < 0.012$ ).

Since the Basic Model did not take any performance measures into account, a second model (called the *Basic Plus Performance Model*) was developed that included performance measures in addition to the variables used in the Basic Model. Productivity measures included clinical income, clinical RVU's, and grant income. Although the pilot study indicated that performance measures of the type considered were not important in the pilot department, the Basic Plus Performance Model served to assess whether performance measures readily available were uniformly unimportant for all departments or whether their importance varied across departments.

The Basic Plus Performance Model was developed using those same basic variables as well as three performance measures, yielding a 17-variable (DF) model that explained 86% of the variability in compensation levels ( $R^2 = 0.864$ ). The pilot study finding of the lack of importance of measures of productivity was validated for the "pilot department." However, the performance measures turned out to be quite important in predicting faculty compensation levels in most other departments.

Using this model to predict compensation, the evidence for gender differences in the deviations between predicted and actual compensation was not formally statistically significant as was found using the Basic Model. However, the gender differences were borderline significant for the entire medical school ( $P < 0.077$ ), with women being paid less than men. In addition, the borderline-significance was accounted for almost entirely by gender differences in the full professors on the investigator track ( $P < 0.075$ ), as was observed using the Basic Model.

Since the Basic Plus Performance Model provides a substantially better fit and considered a more complete set of predictors of compensation, this report is based largely on this model. However, the report begins with an overview of the Basic Model and its findings.

## **The Basic Model**

The School of Medicine GPE Committee met in March 2000 to discuss what data set should be used for the analysis, what variables should be incorporated in the model, and which groups of faculty members should be included in the study. It was agreed that the study should be performed on the most recent academic year data, which was for FY2000. Only the logarithm of the total compensation ( $X+Y+Z$ ) would be analyzed. Variables would include: rank, track, degree, years at Washington University, years since degree, division/department, and whether an individual was a PI on a funded grant. All faculty at the level of assistant professor and above on all three tracks (clinician track, investigator track, and the research track) would be included in the study. Part-time faculty would not be included in the study. Department heads and division directors would also be excluded from the analysis. A total of 945 faculty members was thus included in the study, 731 men and 214 women. The methodology is documented in the **Appendix**.

All departments and divisions were classified into seven department groups based on empirical evidence (see the **Appendix**). Regression analysis was performed predicting the logarithm of total compensation from basic variables and important interactions among them,

keeping only the significant variables or those considered important for the purposes of modeling compensation. This yielded the Basic Model involving 22 variables in all (DF), explaining 77% of the variability in compensation levels at the Washington University School of Medicine ( $R^2 = 0.766$ ).

Under this 22 DF model, we tested for gender differences in the “deviations” (actual compensation minus compensation predicted from the model). Using the Van der Waerden (VW) means test, we compared the group means of the deviations for men against women (the procedure first ranks the deviations within each gender, transforms into Z scores, and then compares the two gender-specific mean Z scores using the t-test). Results are presented in Table I. For the entire group of 214 women and 731 men, the overall medical school wide test yielded  $P = 0.006$ . To explore the source of this difference, we inspected the actual and predicted compensation levels for faculty in each of the track-rank groups under the Basic Model that did not include the gender term (for this purpose, because of small numbers, associate and full professors were pooled together on the clinician and research tracks). As can be seen from Table I, the overall gender difference identified above appears to be largely confined to full professors on the investigator track ( $P = 0.012$ ). In this group, the median percent deviation was -10.67% for women while it is 1.17% for men. The results indicated that, overall, there was a statistically significant difference in compensation by gender. Further investigation suggested that the gender differences were largely confined to the full professor rank on the investigator track, with women paid less than men. A draft report based on this model alone was prepared (version 4 dated June 7, 2001).

**Table I**

**Evidence for Gender Differences in Track and Rank Groups under the Basic Model**

Track	Rank	N <sub>F</sub>	N <sub>M</sub>	Median % deviation (F/M)	P Value
<b>ALL</b>	<b>ALL</b>	<b>214</b>	<b>731</b>	<b>-4.17 / 0.11</b>	<b>0.006</b>
CLINICIAN	ASST	60	93	-0.19 / -0.80	0.106
CLINICIAN	ASSOC/FULL	12	86	-5.84 / -2.58	0.512
INVESTIGATOR	ASST	48	144	-6.22 / 0.34	0.333
INVESTIGATOR	ASSOC	25	138	-1.89 / -0.18	0.130
<b>INVESTIGATOR</b>	<b>FULL</b>	<b>26</b>	<b>194</b>	<b>-10.67 / 1.17</b>	<b>0.012</b>
RESEARCH	ASST	32	56	5.75 / 0.25	0.397
RESEARCH	ASSOC/FULL	11	20	1.56 / 1.31	0.460

**The Basic Plus Performance Model**

Under this model, all departments and divisions were grouped into four Department/Division groups based on the relationship between compensation and all predictive variables including measures of performance (see the **Appendix**). Measures of performance were: clinical income; clinical RVU's, and grant income. Regression analysis, following the

same methodology, yielded the Basic Plus Performance Model involving only 17 variables (DF) with an  $R^2 = 0.864$ . The model is presented in Table II. Note that the coefficients are on the log scale. As can be seen, most variables are highly significant.

Under this 17 DF model, we tested for gender differences in the “deviations” (residuals) from the model. Using the Van der Waerden (VW) means test, we compared the group means of the deviations for men against women. Results are presented in Table III. For the entire group of 214 women and 731 men, the overall medical school wide test yielded  $P = 0.077$ . Because of the borderline nature of this P value, we also investigated whether gender differences might exist within any of the track-rank groups. We compared the group means

**Table II**  
**Definition of the 17 Variables in the Basic Plus Performance Model (including performance) with 17 DF and  $R^2 = 0.864$**

Source		Coefficient (normalized)	DF	P -Value
<b>Med School-Wide Terms:</b>				
Intercept		100%	1	<0.0001
Track	CLIN INVEST	41% 31%	2	<0.0001 <0.0001
Rank	PROF ASSOC	40% 23%	2	<0.0001 <0.0001
Pure MD Degree		40%	1	<0.0001
RVU (per 5,000)		8%	1	0.0004
Grants (per \$1 M)		9%	1	0.0013
Recent Degree (last 3 yrs)		15%	1	0.0202
Full-Prof PIs		14%	1	0.0001
MD/PhD on CLIN or INV Track		24%	1	<0.0001
<b>Additional Department-Group-Specific Terms:</b>				
Groups	1 2 4	49% 21% - 7%	3	<0.0001 <0.0001 0.0041
RVU (per 5,000)	Group 2 Group 3	30% 24%	2	<0.0001 <0.0001
Clinical Revenue (per \$0.5M)	Group 1	42%	1	<0.0001
NO Clinical Revenue	Groups 1,2	- 17%	1	<0.0001

within each of the seven track-rank groups. The smallest P value was observed for the investigator track full professors ( $P = 0.075$ ). In this group, the median percent deviation was -7.29% for women and -0.33% for men. To examine the degree to which this group is

influencing the medical school wide test for gender differences, we deleted this group of faculty and fitted the model again by estimating all coefficients (for the same 17 variables). The overall medical school wide test now yields  $P = 0.305$ , suggesting the investigator track full professors as the major source of possible gender differences at the School of Medicine. In fact, deleting no other track-rank group improved the overall  $P$  value beyond 0.108 (which was obtained when investigator track associate professors were deleted).

**Table III**

**Evaluating gender differences in the deviations from the Basic Plus Performance Model**

Track	Rank	N <sub>F</sub>	N <sub>M</sub>	Median % deviation (F/M)	P Value
<b>ALL</b>	<b>ALL</b>	<b>214</b>	<b>731</b>	<b>-3.00 / -0.27</b>	<b>0.077</b>
CLINICIAN	ASST	60	93	-0.16 / 0.55	0.692
CLINICIAN	ASSOC/FULL	12	86	-2.18 / -2.11	0.576
INVESTIGATOR	ASST	48	144	-4.34 / 0.57	0.483
INVESTIGATOR	ASSOC	25	138	-4.16 / -0.79	0.465
<b>INVESTIGATOR</b>	<b>FULL</b>	<b>26</b>	<b>194</b>	<b>-7.29 / -0.33</b>	<b>0.075</b>
RESEARCH	ASST	32	56	-2.72 / -1.49	0.838
RESEARCH	ASSOC/FULL	11	20	2.40 / 22.50	0.090

**Discussion and Recommendations**

Gender differences have been found in national pay equity analyses as well as in studies undertaken by specific institutions. In 1999, the Massachusetts Institute of Technology admitted that their female faculty “suffer from pervasive, if unintentional discrimination”[1]. This discrimination took the form of differences in salaries, resources provided and the treatment of women faculty. The MIT report documented historical bias that is “subtle, but pervasive” [2]; over a career there can be “an accumulation of slight disadvantages.” That report describes “differences in salary, space, awards, resources, and response to outside offers between men and women.”

More recently, Ginther [3] used data from the national Survey of Doctorate Recipients to evaluate employment outcomes for women in science and engineering. Analysis of salary differences indicated that over time the differences in male and female salaries at the assistant and associate professor levels can be explained by observable characteristics, including productivity. However, she found that substantial gender salary differences among full professors could not be explained by observable differences. In 1997, this amounted to a 15% salary gap overall for women in science and engineering but a 23% gap for women full professors at medical schools. She concluded that nationwide, “gender discrimination similar to

that observed at the Massachusetts Institute of Technology accounts for unexplained gender disparities.”

As a result of the pay equity study carried out at the School of Medicine in 1990, several recommendations were made by the Pay Equity Committee and adopted by the Executive Faculty. These recommendations included immediate review of the salaries of women faculty and the establishment of criteria by which to evaluate remuneration. It was also recommended that a standing pay equity committee should review pay equity data on an annual basis. Women faculty's salaries were adjusted in 1990 and the XYZ compensation plans were developed in 1998. However, no statistically-based analyses of compensation were performed between 1993 and the current pay equity study.

The Basic Plus Performance Model using all the data for FY2000 indicates that the evidence for gender differences in faculty compensation at the School of Medicine, though not formally statistically significant, is of borderline ( $P = 0.077$ ) significance with women paid less than men. Gender differences among the investigator track full professors ( $P = 0.075$ ) appear to be the primary source of the overall evidence of gender differences in the study. Indeed, when the investigator track full professors are excluded from the analysis, the evidence for medical school wide gender differences disappears ( $P = 0.305$ ). In this group of full professors on the investigator track, the median percent deviation was -7.29% for women and -0.33% for men.

Medical school wide, the median percent deviation for female faculty was -5.72% in FY1990 (see **Appendix**) and -4.17% in FY 2000, using the Basic Model to analyze both data sets. Thus, some progress has apparently been made, in that the overall differences among men and women have been reduced. Nonetheless, there is some cause for concern because potential gender differences among investigator track full professors continue to be suggested. In 1990, the median percent deviation for female full professors was -10.41% using the Basic Model compared to -10.67% in a similar analysis of FY2000. Thus, little progress appears to have been made in this particular group. The data do not provide any insight into the reasons for this potential gender gap in compensation. Reasons could include differences in measures of productivity not included in this analysis (such as publications), market factors or gender bias. The committee recognized that certain important activities such as teaching, to which the institution is committed and which are broadly encouraged, have not been taken into account.

While the Basic Plus Performance Model, which includes performance measures, reveals no formally statistically significant gender differences at the medical school or in any of the track-rank groups (see Table III), the committee concludes that the borderline nature of the  $P$  values for the entire medical school as well as for the full professors on the investigator track suggests that one can not dismiss the existence of some gender differences. Moreover, although two-sided tests are justified here, using one-sided tests, as discussed in the **Appendix**, will render some of the tests formally significant. In view of a pattern of possible gender differences in faculty compensation at WUSM, whether formally significant or only borderline significant, the committee makes the following recommendations.

1. The Dean will provide each Department Head with a listing of all faculty in their departments, along with their predicted compensation levels and the 95% confidence limits. The Department Heads should use this information to determine whether FY2000 compensation is fair and meaningful, paying particular attention to those female full professors on the investigator track who fall below the 95% confidence range.
2. The Department Heads should inform the Dean in writing of their evaluations of the compensation of each female full professor on the investigator track, indicating any

adjustments made or explaining why no change is warranted. The Dean, in consultation with the Chair of the GPE Committee, will report the results to the GPE Committee. The GPE Committee will carefully review the results and actions, and recommend to the Dean whether further action needs to be taken.

3. Beginning with data for FY2003, the Associate Dean for Faculty Affairs/Office of Faculty Affairs should initiate a biennial analysis of pay equity, to determine whether progress has been made in eradicating the suggestion of any gender differences in compensation. These biennial analyses will use the model that included measures of performance (as in the Basic Plus Performance Model) unless additional information on variables relevant to compensation becomes available. These analyses should be carried out by an outside individual or organization (who are sufficiently familiar with the medical school setup and compensation policies). The results of the biennial analyses will be reported to the Dean. The Dean will report the findings to the Academic Affairs Committee of the Executive Faculty as well as the GPE Committee. Review and adjustment of the compensation levels of female faculty members should continue until two sequential studies provide no suggestion of any gender gap in compensation. Subsequently, pay equity studies should be carried out every five years at the School of Medicine to ensure that pay equity is maintained.

## References

1. Goldberg, Carey. (1999) "MIT Acknowledges Bias Against Female Professors." *The New York Times*. (March 23, 1999): p.1.
2. "A Study on the Status of Women Faculty in Science at MIT," Massachusetts Institute of Technology, 1999.
3. Ginther, D.K. (2001) "Does Science Discriminate Against Women? Evidence from Academia, 1973-1997". Federal Reserve Bank of Atlanta. Working Paper Series. Working Paper 2001-02. February 2001.

# Appendix

## Statistical Approach

Because there are considerably fewer women than men faculty in the School of Medicine, the committee decided to approach the question of gender differences in compensation levels in a two-step fashion. In step one, we fit a **gender-neutral** statistical model of compensation based upon objective measures, such as seniority, rank, track, and productivity, using all faculty of both genders combined. The goal here is to find the most parsimonious model, which explains the greatest proportion of the variance in compensation levels ( $R^2$ ) using the fewest number of predictive variables (degrees of freedom). In the second step, we use the model derived in step one to test for gender differences. Specifically, we examine the deviations between the model-predicted compensation and the actual compensation (residuals). Faculty who are paid less than predicted by the model will have negative deviations, while faculty who are paid more than the model would predict will have positive deviations. A statistically significant difference in the average deviations between the two genders would indicate evidence of gender differences.

Two predictive models were developed, differing primarily in the number and types of predictive variables considered. The first model, the *Basic (B) Model*, considered only measures of seniority, rank, track, specialty, and the type of department as predictors of compensation. This model is similar to the one developed for examining the question of gender differences at the other WU (Hilltop) schools. The second model, the *Basic Plus Performance (B+P) Model*, considered as possible predictors all of the same factors as in the above model, but also considered individual measures of performance, such as grant funding, clinical revenue generated, and Relative Value Units (RVU—a measure of clinical performance used by the US government to reimburse Medicare costs nationwide; Federal Register, 1999).

## Technical Details:

**Response Variable:**  $\log_{10}[\text{Total Compensation} = (X + Y + Z) \text{ for FY 2000}]$

Because, as of FY2000, not all departments had fully implemented the complete XYZ compensation plan, it was decided that total compensation was the only meaningful measure to attempt to predict for the entire School of Medicine for the FY2000. The distribution of annual compensation is well known not to be normal (Gaussian) except in very narrowly defined subgroups, and this holds true for the medical school as a whole. Taking the  $\log_{10}$  transformation is a common way to deal with the long upper tails of such distributions, and has the property that it preserves the order of compensation, so that hypotheses tested on the log scale are logically the same hypotheses as the corresponding ones on the natural scale. As expected,  $\log_{10}[\text{Total Compensation}]$  is sufficiently nearly normally distributed in the medical school to justify the use of regression methods to meaningfully model compensation.

**Data: FY 2000**

### **Inclusion Criteria:**

1. Full-time WUMS faculty who worked the entire year
2. Clinician, Research and Investigator Tracks
3. Assistant, Associate and Full Professors Ranks

**Exclusion Criteria:**

1. Department or Division Heads

This yields 214 female and 731 male = 945 total faculty to be analyzed.

**Predictive Factors Considered:**

**Basic Model:**

- Track (Clinician, Research, Investigator)
  - Rank (Assistant, Associate, Full)
  - Degree (MD, MD/PhD, PhD/other)
  - Years at Washington University
  - Years since terminal degree
  
  - Primary Appointment  
Department/Division\*
- 

**Basic Plus Performance Model:**

- All factors from the Basic Model

**PLUS**

- Clinical Revenue (Total \$ collected)
- RVU (Total RVUs credited)
- Grant Revenue (Total \$ credited;  
Includes PIs of all individual programs in PPGs)

## Department/Division Groups

Since subspecialty makes a huge difference in compensation at many institutions in the field of medicine, it is not surprising that important differences exist in the average compensation between many Department/Divisions in the WU School of Medicine. However, there are FAR too many Departments/Divisions to allow for a separate regression parameter (intercept) to be estimated for each such unit. Some pooling of Departments/Divisions must be done into larger subgroups for the sake of parsimony. While the same principles were applied in defining Department/Division grouping for the two models, the exact way this was done differed in the two models, reflecting the different information used to predict compensation. For the B Model, Department/Divisions were grouped together if they had similar average compensation. By contrast, for the B+P Model, Department/Divisions were grouped together based upon similarity of the degree to which performance measures predicted compensation. In both cases, we first pooled those divisions with fewer than 10 faculty into like groups (according to similar average compensation), so that all units considered would have at least 10 observations upon which to model. For the B Model, we defined seven Department/Division groups. For the B+P Model, we needed only four Department/Division groups.

When considering performance, differences were noticed in the relationship between compensation and income/RVUs even within the Departments of Internal Medicine and Radiology (by Division). We therefore began by dividing Internal Medicine into 12 Divisions and Radiology into four Divisions. In all, now there are 35 “Departments or Divisions,” which are going to be re-grouped according to similarity of the relationship between performance measures and pay. To do the grouping, we developed a new baseline model in which basic variables like track, rank, and degree as well as performance variables like clinical revenue, RVU, and grant funding were all forced into a single model (14 variables in all). In addition, a separate intercept for each of the 35 Departments/Divisions was allowed to enter the model in a step-wise fashion if significant even at a liberal significance level of 0.10. In the end, the stepwise procedure selected 20 of the 35 Department/Division variables, to a total of 34 DF with an  $R^2 = 0.843$ . The remaining 15 Departments/Divisions did not need separate intercepts (and were part of the overall common intercept).

Careful inspection of the additional intercepts for the 20 Department/Divisions suggested that some of the 20 additional intercepts hardly contributed to the fit of the model and therefore could be merged with the overall common intercept (combined with the other 15 Departments/Divisions who did not need separate intercepts, thus saving additional DF). Moreover, even the remaining ones could be pooled into three department groups without compromising on the fit of the model noticeably, thus reducing a total of 35 Departments/Divisions to a total of four department groups (including those merged into the overall intercept). Group 3 was the overall “average” group that included all pre-clinical departments as well as a few clinical Departments/Divisions. Group 4 was below (with a negative additional intercept), and groups 1 and 2 were above (with positive and different additional intercepts)

### Departmental Groups under the Basic Model

Group	Departments/Divisions
1	Dental School, Pediatrics, Neurology
2	Anatomy and Neurobiology, Biochemistry & Molecular Biophysics, Internal Medicine, Pathology & Immunology, Cell Biology & Physiology, Psychiatry, Physical Therapy Program

3	Genetics, Molecular Microbiology, Otolaryngology, Molecular Biology & Pharmacology, Biostatistics, Health Administration
4	Anesthesiology, Neurological Surgery
5	Radiology, Orthopaedic Surgery
6	Obstetrics and Gynecology, Ophthalmology & Visual Sciences
7	Surgery

### **Departmental Groups under the Basic Plus Performance Model**

<b>Group</b>	<b>Departments/Divisions</b>
1	Cardiology Consult Group of Internal Medicine Radiology and Radiation Oncology Divisions of Radiology Orthopaedic Surgery Surgery Anesthesiology
2	Internal Medicine, Cardiology, Emergency Medicine Divisions of Internal Medicine Neurological Surgery Radiation Sciences and Radiology Other Division of Radiology Obstetrics and Gynecology Ophthalmology & Visual Sciences
3	(all other Departments/Divisions)
4	Hematology and Infectious Diseases Divisions of Internal Medicine Pediatrics Neurology

#### **Step 1: Gender-Neutral Predictive Model:**

Regression (PROC REG) and ANOVA (PROC GLM) in SAS™ (SAS Institute, 1989). These two programs give identical answers, since the underlying statistical models are identical.

Selection of predictive variables based upon:

- Stepwise regression (forward with backward elimination of significant predictors)
- Statistical judgment based upon statistical significance and a balance between R<sup>2</sup> improvement and degrees of freedom.

#### **Step 2: Test of Gender Differences in Compensation**

Calculate for EACH individual faculty: Deviation =  $[10^{(\text{Predicted } \log_{10} [\text{comp}] \text{ from Model})} - \text{Actual Compensation}]$

#### **Primary Hypothesis:**

H0: Male Deviation = Female Deviation for entire medical school

Test: Two-tailed van der Waerden non-parametric test (normal scores) of Male vs. Female deviations in all 945 faculty

### **Secondary Hypotheses:**

- H0: Male Deviation = Female Deviation in particular Track/Rank combination (e.g. Full Professor Investigators; Assistant Professor Clinicians, etc.)
- Test: Two-tailed van der Waerden non-parametric test (normal scores) of Male vs. Female deviations in that Track/Rank subgroup [Note: in both the Clinician and Research track, there were insufficient numbers of women to test the Full Professor rank as a separate group (N = 1 and N = 2, respectively). Thus, for each of those two tracks, the secondary tests were done combining the Associate with Full Professor Ranks]

**One-sided versus two-sided:** All analyses and this report, just as the draft report (version 4 dated June 7, 2001 based on the Basic Model alone), are concerned with the question “Were male and female faculty compensated equally in FY2000?”. Therefore, all P values were derived using what are called “two-sided” tests (where the alternative to “*equal compensation*” is “*unequal compensation*”). One may argue that the alternative should be “*are women compensated less than men?*”, especially because there has been a pattern over the years suggesting that women may have been compensated less. Under such an argument, all P values reported would need to be halved, which would render some of them formally significant. In either case, it is clear that some of the tests, especially for the medical school wide test and the investigator track full professors, provide “borderline” evidence suggesting gender differences (whether the “borderline” is on the formal significance side or on the non-significance side).

**Analysis of FY1990 salaries:** Prior to the current analysis, the most recent pay equity study at Washington University School of Medicine was performed in 1990. Because the methodology used in that study was different from that used in the current study, the FY1990 data were re-analyzed using the present methodology. For this analysis, the 19 departments (in FY1990) were pooled into five department-groups, rather than into seven groups as had been done for the FY2000 data. The fit of the model to the FY1990 data that included separate variables for the five department groups was excellent, with an  $R^2 = 0.767$ . This, in fact, provides a better fit of the data than the model used in 1990. Consistent with what was found in the original study, analysis of the FY1990 data using the current methodology indicates the presence of a significant gender gap in faculty salaries in 1990. Medical school wide, the estimated median percent deviation for women was -5.72% for women (i.e., women were paid 5.72% less than that predicted). Unlike the situation with the FY2000 data, analysis of the FY1990 data by track and rank failed to identify any particular subgroup where the gender differences were concentrated.

Finally, it should be remembered that because the most parsimonious prediction model was based on the entire data involving all 945 faculty, separate tests within track/rank groups using that same model are correlated to a degree. In particular, any action involving one group (such as the investigator track full professors) can have an impact on the other tests (such as, for example, the research track associate/full professors with the next smallest current P value of 0.090).

### **References:**

- Federal Register: November 2, 1999 (Volume 64, Number 211) Rules and Regulations, From the Federal Register Online via GPO Access [wais.access.gpo.gov][DOCID:fr02no99-16] Pages 59379-59428
- SAS Institute Inc. (1989) SAS/STAT User's Guide, Version 6, Fourth Edition Volume 1 & 2, Cary, NC, SAS Institute Inc.