ELSEVIER

Journal of Economic Behavior & Organization
Vol. 63 (2007) 514–536

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

Pushing incomes to reference points: Why do male doctors earn more?

John A. Rizzo[a,b,*], Richard J. Zeckhauser[c]

[a] *Department of Economics, Stony Brook University, United States*
[b] *Department of Preventive Medicine, Stony Brook University, United States*
[c] *John F. Kennedy School of Government, Harvard University, United States*

Received 4 February 2004; received in revised form 7 February 2005; accepted 2 June 2005
Available online 5 May 2006

Abstract

Why do female physicians earn less than their male counterparts? Data from the Young Physicians Survey yields an answer. Young male and female physicians respond differently to the reference incomes (RIs) they state. Males (revealing prospect-theoretic preferences) respond strongly, particularly if below RIs. Females do not respond. Males also set higher RIs. That, combined with males' greater responsiveness to RIs, fully explains the significant gender gap in earnings and earnings growth rates. Together, productivity measures and prejudice play at most a modest role. To boost their incomes, males raise their hourly incomes, not their hours of work.

JEL classification: I11; J31

Keywords: Reference incomes; Male doctors; Behavior

1. Introduction

Female physicians earn less than their male counterparts.[1] Why? Possible explanations include differences in productivity as indicated by human capital measures (such as specialty mix or

* Corresponding author at: HSC, Level 3, Rm 071, Stony Brook University, Stony Brook, NY 11794-8338, United States. Fax: +1 631 444 3480.
E-mail address: john.rizzo@stonybrook.edu (J.A. Rizzo).

[1] Bashaw and Heywood (2001). Early studies (Kehrer, 1976; Bobula, 1980; Langwell, 1982; Mitchell, 1982) reported a substantial earnings gap, with female physicians earning 20–25 percent less than males after adjusting for differences in demographic and medical practice characteristics. Culler and Ohsfeldt (1986) report a smaller though still substantial unexplained gap of 12–13 percent. These authors attribute their smaller estimated difference to more fully specified

0167-2681/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2005.06.007

board certification), differences in preferences (say for practice style, specialty, or hours of work), or prejudice by employers or patients. Empirical studies have typically employed a standard decomposition technique to determine the extent to which gender-specific earnings differences reflect observable factors thought to be correlated with productivity (often measured as a human capital term), and they also identify a residual or unexplained effect.[2] This unexplained residual is often interpreted to reflect gender discrimination (for example, patients' reluctance to go to women, and physicians' reluctance to refer to them). Unfortunately, previous studies have lacked direct information on gender prejudice, making it difficult to assess whether discrimination is really present and what it implies about earnings.

Factors other than discrimination may contribute to the residual; for example, it may reflect important gender-specific differences in how men and women respond to the same incentives, implying differences in preferences toward income or in activities to produce it, or both.[3] If so, further study is required to disentangle the competing explanations for the residuals.

Past research has documented differences in the preferences of male and female physicians. For example, female physician labor supply is more elastic with respect to hourly earnings opportunities than is male physician labor supply (Rizzo and Blumenthal, 1994).[4] Clinical evidence suggests that female physicians typically choose less invasive procedures and favor more involvement by their patients in treatment decisions than do their male counterparts.[5]

No previous research has considered whether these or other gender-specific differences in physicians' behavior might lead to gender-specific differences in earnings. The literature has assumed, often implicitly, that differences in male and female physicians' preferences affect their earnings only to the extent that those differences are reflected in human capital or other endowments. No study has tested the hypothesis that male and female physicians' income-generating activities, such as hours worked, choice of practice setting, or patients seen per hour, differ systematically. We focus here on reference income behavior, exploring the hypothesis that males and females set different standards for adequate incomes and respond to these targets quite differently when their actual incomes fall short of their reference levels.

We draw upon behavioral economics to understand how reference income influences income generation. The critical constructs are from prospect theory (Kahneman and Tversky, 1979), which asserts that reference points frame decisions, thereby influencing behavior and loss aversion.[6] In earlier research (2003) we found that young male physicians' self-designated reference incomes were an important determinant of their income.

Doctors who set challenging income objectives may become less faithful agents for their patients, say by reducing their time per visit or refusing to treat uninsured patients. If female

---

earnings equations and "more accurate procedures to calculate the unexplained earnings differential from the estimated earnings equations" (p. 345). Using the same specification and method for calculating the unexplained differential as in previous studies, Culler and Ohsfeldt find an earnings differential of 20 percent, which closely agreed with the earlier studies. Their new specification (including expanded measures of practice experience and physician specialty) reduced this differential to 15 percent. Corrections recognizing that the dependent variable (the natural logarithm of earnings) is log-normally distributed further reduced the differential to 12–13 percent.

[2] See, for example, Blinder (1973) and Oaxaca (1973).

[3] See Wood et al. (1993).

[4] This finding has also been observed for a number of other industries. See Killingsworth (1983) for further details.

[5] See, for example, Bickell et al. (1994), Britt et al. (1996), Domenighetti et al. (1985), Lurie et al. (1993), Mitler et al. (2000), Wertz (1993).

[6] Loss aversion is a central element of prospect theory, which deals primarily with choice under risk. However, risk is not a central element of our analysis.

physicians care about their patients' well-being more than male doctors do, then reference incomes will affect their practices less and will have a smaller effect on income growth.

The present study focuses on differences in behavior motivated by reference incomes (that is, differences in preferences), as an explanation for gender differences in income. First, we test whether men and women respond differently when income falls short of reference income. Second, we test whether men set higher reference incomes. Third, we examine whether such differences significantly explain the gender-earnings gap. Fourth, we briefly test for two alternative explanations for this gap, namely productivity differences and discrimination.

We use a unique panel of physician data, the Young Physicians Survey (YPS). Hence, we examine only the behaviors of young doctors. We exploit the panel nature of the data, comparing physician-specific differences in earnings growth over time. That is, our multivariate estimates focus on earnings *growth* rather than on cross-sectional differences in earnings *levels*.[7] We believe that this is a useful approach. First, it allows us to examine how reference incomes affect physician-specific differences in earnings over time. Second, because we focus on young physicians in their practice-building years, earnings growth is substantial, providing enough variability to yield statistical lessons.[8] We also use the multivariate estimates to generate predicted absolute earnings levels, allowing us to investigate whether and to what extent gender differences in earnings levels reflect gender differences in reference income behavior, productivity, or prejudice.

Section 1 of this paper describes the data used in our analysis and provides descriptive evidence on earnings and reference incomes for male and female physicians. Evidence on gender-specific differences in behavior from the psychology, economics, and medical literatures is discussed in Section 2. Section 3 presents an earnings shortfall model that incorporates the notion of reference incomes. Hypotheses, empirical models, and econometric issues are discussed in Section 4. Section 5 presents results on the reference income model, and on productivity and prejudice. Section 6 summarizes the findings and their implications.

2. Data and descriptive evidence

2.1. Data

We use data from the 1987 and 1991 Practice Patterns of Young Physicians Surveys (YPS).[9] These telephone surveys were designed to provide information for a nationally representative sample of young physicians practicing in the United States during 1987 and 1991. The YPS surveys include information on the respondents' demographic characteristics, medical practice patterns, family background, and current income and expenses. Respondents were also asked about their medical specialties, employment arrangements (e.g., HMO, private practice), number of physicians in practice, and so on.

The physicians' reference income plays a critical role in our model. It was defined as follows. The YPS asked physicians: "Considering your career stage, what do you consider to be an adequate income after expenses but before taxes from your professional activities?" We take the response to this question to indicate reference income. The YPS also asked physicians about their incomes during the previous year, after practice-related expenses but before taxes.

[7] Due to our logarithmic specification, we examine differences in geometric mean values of earnings growth.

[8] On average, income grew by about 80 percent between 1986 and 1990 for the physicians in our sample (see Table 2).

[9] The YPS surveys were conducted by the American Medical Association under a grant from the Robert Wood Johnson Foundation.

"Young" physicians were defined as physicians under the age of 40 in 1987 who practiced for 1–8 years since their residencies. A random sample 50 percent of the 1987 survey respondents was resurveyed in 1991. Each survey was conducted by telephone and achieved response rates of about 70 percent. We restricted our sample to self-employed physicians. Such physicians (the majority of doctors practicing in the United States) can better alter their practice patterns to achieve reference incomes than employee physicians, who typically receive a salary. Radiologists, pathologists, and anesthesiologists were also excluded, as these specialists tend to be hospital-based and have less discretion in setting their practice patterns.[10]

These exclusions left a sample of 735 physicians for whom we had data in both 1987 and 1991. Of these physicians, 464 had incomes below their reference points and 271 had incomes at or above their reference points. Roughly 85 percent of each cohort was male. Prior work (Rizzo and Zeckhauser, hereafter (2003)) showed that reference incomes had no effect for male physicians who earned more than their reference incomes. To determine the explanatory power of reference incomes, as well as to ease exposition, we focus here on physicians whose incomes are below their reference points. Since selecting a sample in this way could conceivably introduce a bias, we also provide a parallel study using the entire sample (physicians both above and below their reference incomes) in Appendix B.[11,12] The results are nearly unchanged; in particular, in the below-reference-income cohort, reference income behavior significantly affects income for males but not females. The names, descriptions, means, and standard deviations of the variables used are provided in Table 1.

2.2. *Physician earnings and reference incomes*

Male physicians earn substantially more than females, and they set higher reference incomes. They have modestly higher growth rates in actual and in reference incomes. Table 2 provides data for the roughly 60 percent of physicians below reference incomes in 1986–1987, who are our focus, as well as for all physicians.

Table 3 presents descriptive results on the relationship between reference income and earnings growth. The table shows the mean values of the ratio of 1987 reference to 1986 actual income (REFRATIO) and the ratio of 1990 to 1986 income (INCRATIO), 1990 to 1986 hourly wages (WAGERATIO), and 1991 to 1987 hours worked (HOURATIO). Correlations between REFRATIO and INCRATIO, WAGERATIO, and HOURATIO are provided as well. Turning first to the below-reference-income cohort, we see that, for both males and females, reference incomes

[10] To remove undue influence of statistical outliers, we omitted the highest and lowest 1 percent of male and female earners for 1986 and 1990. A doctor whose income was clearly mis-transcribed was also excluded. The results were quite similar to those without any excluded observations. We also obtained qualitatively similar results when we excluded the top and bottom 5 percent of income measures for males and females. Our findings are not the artifact of a few extreme values.

[11] Appendix B is available on the JEBO website (www.elsevier.com/locate/econbase).

[12] The models in Appendix B are more complex than those in the text because they include interaction terms that allow the effects of reference income to differ according to whether physicians are above or below their reference incomes and also by gender. Not included, but available on request, are results demonstrating that reference incomes have small and statistically insignificant effects on income growth when the sample is confined to at-or-above-reference-income physicians. The top and bottom 1 percent of male earners in each period were excluded from the above-reference-income cohort to remove the influence of extreme outliers. No females were excluded, since 1 percent of the above-reference-income sample of 41 female physicians rounded to zero. We also computed the results without these exclusions. They were quite similar to those reported in the Appendix B.

Table 1
Variable names and descriptive statistics

Variable name	Description	Males ($N = 398$) mean (S.D.)	Females ($N = 66$) mean (S.D.)	All ($N = 464$) mean (S.D.)
ln INCOME90	Natural log of physician income ($1000s) in 1990	4.95 (0.54)	4.62 (0.50)	4.90 (0.54)
ln WAGE90	Natural log of hourly wage in 1990	3.89 (0.57)[a]	3.66 (0.46)[a]	3.85 (0.56)[b]
ln HOURS91	Natural log of annual hours worked in 1991	7.99 (0.27)[a]	7.87 (0.29)[a]	7.97 (0.28)[a]
ln WAGE86	Natural log of hourly wage in 1986	3.19 (0.54)[a]	3.14 (0.55)[a]	3.18 (0.54)[a]
ln HOURS87	Natural log of annual hours worked in 1987	7.99 (0.29)[a]	7.84 (0.41)[a]	7.96 (0.32)[a]
ln REFINC87	Natural log of reference income ($1000s) in 1987	4.68 (0.41)	4.42 (0.44)	4.64 (0.43)
ln INCOME86	Natural log of physician income ($1000s) in 1986	4.26 (0.47)	4.04 (0.45)	4.23 (0.47)
PRACSIZE	Number of physicians in practice	10.26 (29.37)	5.62 (13.27)	9.60 (27.69)
PRACYEAR	Years of practice experience	2.71 (1.37)	2.36 (1.40)	2.66 (1.38)
GENIM[b]	Dummy variable (DV) = 1 if physician specialty is general internal medicine, otherwise = 0	0.15 (0.36)	0.15 (0.36)	0.15 (0.36)
PED[b]	DV = 1 if pediatrician	0.10 (0.30)	0.30 (0.46)	0.13 (0.34)
SPECIM[b]	DV = 1 if internal medicine with subspecialty	0.15 (0.36)	0.06 (0.24)	0.14 (0.35)
GENSURG[b]	DV = 1 if general surgeon	0.06 (0.24)	0.00 (0.00)	0.05 (0.22)
SPECSURG[b]	DV = 1 if specialized surgeon	0.20 (0.40)	0.09 (0.29)	0.18 (0.39)
PSYCH[b]	DV = 1 if psychiatrist	0.04 (0.20)	0.05 (0.21)	0.04 (0.20)
OBGYN[b]	DV = 1 if obstetrician/gynecologist	0.09 (0.28)	0.17 (0.38)	0.10 (0.30)
FEMALE	DV = 1 if physician is female	–	–	0.14 (0.35)
MARRIED	DV = 1 if physician is married	0.87 (0.34)	0.80 (0.40)	0.86 (0.35)
NUMKIDS	Number of children physician has	1.46 (1.29)	1.09 (0.96)	1.41 (1.25)
BLACK	DV = 1 if black	0.10 (0.30)	0.18 (0.39)	0.11 (0.32)
BDCERT	DV = 1 if board certified	0.70 (0.46)	0.61 (0.49)	0.69 (0.46)
FMG	DV = 1 if graduate of foreign medical school	0.10 (0.30)	0.08 (0.27)	0.09 (0.29)
RURAL[c]	DV = 1 if rural practice	0.05 (0.21)	0.03 (0.17)	0.05 (0.21)
NONSMSA[c]	DV = 1 if located in non-rural non-SMSA	0.15 (0.36)	0.14 (0.35)	0.15 (0.36)
SMALSMSA[c]	DV = 1 if located in small SMSA	0.20 (0.40)	0.15 (0.36)	0.20 (0.40)
BIGSMSA[c]	DV = 1 if located in large SMSA	0.14 (0.34)	0.12 (0.33)	0.13 (0.34)
MIDWEST[d]	DV = 1 if located in Mid West Census Region	0.17 (0.38)	0.21 (0.41)	0.18 (0.38)
SOUTH[d]	DV = 1 if located in South Census Region	0.34 (0.47)	0.33 (0.48)	0.34 (0.47)
WEST[d]	DV = 1 if located in West Census Region	0.23 (0.42)	0.29 (0.46)	0.24 (0.43)

Note: Unless otherwise specified, all data are as of 1987.

[a] Sample sizes are slightly smaller for these variables due to missing values in hours worked for a few physicians.

[b] Excluded reference specialty is general/family practice.

[c] Excluded reference location is largest SMSA category (>1 million population).

[d] Excluded reference census region is the Northeast.

Table 2
Mean earnings, reference incomes, and growth rates by gender

Variable	Males ($N=628$) mean	Females ($N=107$) mean	All ($N=735$) mean
All physicians			
Actual income in 1986 ($1000s)	104***	79***	101
Actual income in 1990 ($1000s)	181***	125***	178
Reference income in 1987 ($1000s)	121***	93***	117
Reference income in 1991 ($1000s)	190***	132***	182
Percent growth in actual income[a]	74	58	76
Percent growth in reference income[a]	56	42	56
Variable	Males ($N=398$) mean	Females ($N=66$) mean	All ($N=464$) mean
Physicians below their reference incomes in 1986–1987			
Actual income in 1986 ($1000s)	79***	63***	76
Actual income in 1990 ($1000s)	164***	115***	157
Reference income in 1987 ($1000s)	117***	92***	114
Reference income in 1991 ($1000s)	181***	130***	173
Percent growth in actual income[a]	108	83	107
Percent growth in reference income[a]	55	41	52

[a] These figures are calculations of differences in mean growth in income for the sample populations, not the mean value of individual growth rates for physicians, thus they are not amenable to statistical testing.
*** Statistically significant difference in this value for males and females at the 1 percent level, two-tailed test.

are substantially above realized income. Consistent with the results in Table 2, income grew substantially in all cases. This rapid income growth is not surprising; our sample consists of young physicians who are building their practices.

More striking is the gender-specific difference in the correlation between the ratio of reference to actual income and the ratio of 1990 to 1986 income (REFRATIO and INCRATIO).

These variables are highly correlated for males (correlation coefficient = 0.62), indicating that male physicians whose reference incomes are above their realized incomes achieve higher earnings growth in subsequent years. This pattern is consistent with what one would expect if such

Table 3
Reference incomes and earnings for physicians below their reference points

	Males ($n=398$)[a]	Females ($n=66$)[a]	All ($n=464$)[a]
REFRATIO (ratio of 87 reference to 86 income)	1.62	1.50	1.60
INCRATIO (ratio of 90 income to 86 income)	2.28	1.97	2.24
WAGERATIO (ratio of 90 hourly wage to 86 hourly wage)	2.40	1.95	2.33
HOURATIO (ratio of 90 hours to 86 hours)	1.05	1.14	1.06
Correlation: REFRATIO:INCRATIO	0.62***,[b]	0.28***,[b]	0.61***
Correlation: REFRATIO:WAGERATIO	0.57**,[c]	0.34**,[c]	0.56***
Correlation: REFRATIO:HOURATIO	−0.01	−0.15	−0.03

[a] Sample sizes are somewhat smaller for correlations involving hourly wages and hours worked due to missing values in hours worked by some respondents. Thus, in the males' sample, wage and hours worked correlations with INCRATIO are based on 354 observations. In the female sample, these correlations are based on 63 observations, and in the full sample, 417 observations.
[b] These correlations are significantly different for males and females at the 1 percent level.
[c] These correlations are significantly different for males and females at the 5 percent level.
** Statistically significant at the 5 percent level, two-tailed test.
*** Statistically significant at the 1 percent level, two-tailed test.

physicians respond strongly to reference incomes. In contrast, the correlation between REFRATIO and INCRATIO for females is less than half as large as for males (correlation coefficient = 0.28). For females, relatively high reference points are substantially less predictive of relatively high rates of earnings growth. The difference in correlation for males and females is significant at the 1 percent level on a two-tailed test.[13] This gender-based difference in response to reference incomes is explored at length in the multivariate analysis below. After controlling for other factors, the relationship for males remains strong. For females it is statistically insignificant.

The strong positive correlation between REFRATIO and INCRATIO for male physicians who have not met their reference incomes is consistent with the notion that reference incomes affect income growth for this group. The weaker or nonexistent role for females suggests that reference incomes play less of a role for them.

The correlations between INCRATIO and WAGERATIO are also significantly stronger for males than for females. In contrast, INCRATIO is unrelated to HOURATIO for both males and females. These results suggest that gender differences in the relationship between reference income and income growth reflect behaviors that raise hourly earnings rather than increases in hours of work. We investigate this issue further using multivariate analysis below, finding results consistent with these patterns.

Finally, for physicians who are at or above their reference points, the relationship between REFRATIO and INCRATIO is much weaker. It is not statistically significant for females. For males, though significant, it is less than one-fourth as great as the correlation for males below their reference incomes.[14] Among males, the dramatic difference between those above and below their reference incomes is predicted by loss aversion, which posits that the utility function is much steeper below than above a reference point. Thus, males will struggle to push income up when below the reference point, but push much less when above it.[15]

2.3. *Do males and females behave differently?*

The hypothesis that male physicians vigorously try to increase their incomes in response to shortfalls from reference incomes while females do not is controversial. However, a substantial body of evidence from economics, medicine, and psychology points to important gender-specific differences in behavior.[16] Behavioral economics has been particularly sensitive to gender-based differences, documenting, for example, that females are more cooperative and trusting (Camerer, 2003).

In a study of gender differences in earnings, Bowlus (1997) reports that "nontrivial portions of the wage differentials are attributable to gender differences in labor market behavior" (pp.

[13] We used the Fisher's z transformation (Kennedy, 1992) to test for differences in these correlations.

[14] Specifically, in the cohort at or above reference income, correlations between INCRATIO and REFRATIO are 0.14 for males (significant at the 5 percent level), 0.07 for females (not statistically significant); and 0.13 overall (significant at the 5 percent level). Correlations between REFRATIO and, respectively, WAGERATIO and HOURATIO are statistically insignificant for both males and females in this cohort.

[15] See (2003) for the earnings case. Bowman et al. (1999) examine the implications of loss aversion for consumption behavior. In an uncertain world, they show that consumers will resist lowering consumption when income falls, but be more relaxed about raising it when income rises. They confirm this hypothesis using aggregate data on five major developed nations.

[16] Some observers would ascribe many of these behavioral differences to discrimination, although not necessarily discrimination in the workplace, our area of concern. For example, if women but not men tend to quit work for childrearing, and if childcare burdens are imposed upon them differentially, women's higher rates of job quits might be attributed to discrimination.

629–630). Wood et al. (1993) study graduates of the University of Michigan Law School (who are highly trained professionals like our physicians) for behavioral differences between the genders. Finding one-third of the gender earnings gap unexplained, they conclude, "Overall, our results provide support for both behavioral and structural explanations of the wage gap" (p. 440). Behavioral economists have studied gender differences in behavior in the more controlled environment of simple laboratory experiments. They have found, for example, that women are less likely to reject offers in ultimatum games (Camerer, 2003, p. 64).

Studies of earnings differentials suggest that men and women have different beliefs about what matters in a job. For example, Daymont and Andrisani (1984) report that 22 percent of males indicate that making money is a very important characteristic of a job, while only 11 percent of women respond this way. Women are substantially more likely to believe that helping others is a very important job characteristic of a job (72 percent) than are men (47 percent).[17]

Comparative analyses of physicians' practice patterns suggest that females may be less willing than their male counterparts to give up other valued attributes for additional income. Female physicians work fewer hours per week and are more likely to interrupt their medical careers (Bowman and Gross, 1986). Female physicians may also be more reluctant to engage in income-generating activities that might adversely affect their patients. One of the most consistent findings in the literature is "that female physicians spend more time with patients than do their male colleagues." Most studies also show that women are better "at involving patients in decisions and explaining medical terminology" (Harvard Health Letter, 2001, p. 4; see also Bowman and Gross, 1986). Perhaps this is because, as psychosocial theorists contend, females work in a more relationship-based structure with their patients, encouraging patient responsibility and choice (Wertz, 1993; Martin et al., 1988). In a similar vein, other researchers (Gilligan, 1982; Johnson, 1988; Lyons, 1988) have contended that women and men tend to develop different moral senses, with women more likely to "value human relationships and feelings and make their moral judgments on the basis of damage to those relationships" (Brannon, 1996, p. 132).[18]

Clinical decisions differ significantly between male and female physicians as well, with male physicians opting more for invasive and costly treatments, such as surgery. For example, male obstetrician–gynecologists perform hysterectomies at twice the rate of their female counterparts (Domenighetti et al., 1985; Bickell et al., 1994), and they perform cesarean sections up to three times as often (Mitler et al., 2000). By contrast, female GPs offer their women patients more

[17] Each of these differences was found to be statistically significant. See Daymont and Andrisani (1984) for further details.

[18] Brannon (1996, p. 133) relates the following story illustrating the gender-specific difference toward personal caring emphasized by Gilligan and colleagues:

> A woman and a man each told me about judgments about right and wrong made by men and women and how odd these judgments seemed to them, as though each was using a different moral sense to make their judgments.. . .The man was a retired supervisor and said that he found working with women difficult because they seemed to take everything personally.. . . The woman talked about her husband's business dealings in much the same terms; he considered things business that she considered personal. "I don't see how a man can do business with someone he considers a sleaze. That just seems wrong to me. How can a person fail to take personal characteristics into account? In a sense, everything is personal – how you treat people – what can be more personal? I would find it very difficult to do business with someone I considered a bad person, and I don't think I could hurt someone I liked in the name of a business deal. My husband will try to hurt people in business deals that he likes on a personal level. It's like a game. That seems wrong." This lesson extrapolates to the medical arena. If male doctors, particularly those with income shortfalls, consider doctoring primarily as a business, whereas female doctors consider it personal, and there is some potential for questionable though profitable activities, then males will earn more.

gender-specific screening, such as mammograms and pap smears, than do their male counterparts (Lurie et al., 1993; Britt et al., 1996). The male tilt toward invasiveness, the female tilt toward prevention persists even after controlling for a variety of patient and medical practice characteristics. This leads some researchers to conclude that such differences are to some extent "inherent to the sex of the physician" (Britt et al., 1996, p. 403).

A final behavioral factor that might explain gender earnings differentials are women's attitudes toward their families. Fuchs (1988) studies the range of posited factors that lead to a gender earnings gap and concludes, "the biggest sources of women's economic disadvantage [is] their greater desire for and concern about children." He cites as support that "the wages of mothers in their thirties . . . lag (relative to childless women) by from 7 to 10 percent for each additional child" (p. 140).[19]

In sum, evidence from economics, medicine, and psychology suggests that differences in the economic and clinical behaviors of male and female physicians should be expected. We turn next to our conceptual framework and an empirical investigation of such gender-specific differences.

3. An earnings shortfall reference income model

Following earlier work (2003), we employ a model of reference income behavior as our conceptual framework.[20] The physician service industry is a good place to find reference income behavior, if it exists. First, physicians have some discretion in setting prices and significant discretion in selecting treatments.[21] Information problems make it difficult for consumers to judge the quality of the services they receive or to shop comparatively, and health insurance makes them relatively insensitive to price. These factors increase physicians' control over treatment decisions, facilitating their ability to push toward desired reference incomes. The idea that reference or target incomes affect physicians' behavior – often referred to as the "target income hypothesis" – has sparked great debate in the health economics literature.[22] Our intention is not to revisit this debate, but to test the hypothesis that the reference income behaviors of physicians differ by gender. Specifically, we test the hypothesis that males experiencing earnings shortfalls respond much more vigorously to reference incomes than do females in this situation.

Before turning to our tests, we sketch the underpinnings of our model. Details are given in (2003).[23] The theory of loss aversion (Kahneman and Tversky, 1979) is a vital component of reference income behavior. Loss aversion posits that *reference levels* are important in determining preferences. The central lesson of loss aversion is that the preference function is steeper in the

[19] Female physicians are in an elite and male-dominated profession. Intuition alone cannot tell us whether Fuchs' findings about child-orientation would extend to them.

[20] Reference income effects have received little attention from contemporary economists; why is unclear. Frank (1985) addresses this question using the term income targets rather than reference income. He argues that economists reject income targets because these targets require that comparisons of relative income matter, whereas economists have been trained to model utility with absolute measures of consumption and income, not with relative measures. See (2003) for further discussion of this issue and references to related works.

[21] Given increased cost control measures in recent years, however, actual fee reimbursement has become increasingly stringent. During the period of this analysis (1987–1991), the self-employed physicians we examine had more discretion in setting prices than do physicians today.

[22] "Target income" is the popular term in the health economics literature. We prefer the term "reference income" because it is more consistent with the terminology of loss aversion.

[23] That analysis also discusses arguments for and against the notion of reference income behavior, particularly in relation to physicians, and includes references to earlier research addressing this topic.

negative direction than in the positive one. A growing literature supports the idea of loss aversion. Researchers have found, for example, that subjects are more sensitive to losses than to foregone gains and to out-of-pocket costs than to opportunity costs (Kahneman and Tversky, 1984; Thaler, 1980). Dunn (1996) finds that workers' indifference curves display sharp bends near their current wage and leisure positions, a labor economics finding consistent with the predictions of loss aversion.[24]

For physicians, reference incomes (what the physician reports as an "adequate income") serve as the reference level, or zero point, for evaluating gains or losses. For example, additional income should be more highly valued among physicians whose incomes are below their reference points than those whose incomes are at or above their reference levels. Indeed, in our earlier study of male physicians, we found that among physicians whose incomes were below their reference levels, higher reference incomes were directly and significantly related to subsequent gains in income. By contrast, reference incomes did not influence the incomes of physicians who were at or above their reference points (2003). Physicians who are eager to close an income shortfall have two choices. They can work more hours or find a way to earn more per hour. Hourly income can be boosted by charging more, cutting time per visit, tilting toward more remunerative procedures, or moving to a more lucrative setting. We label such practices as stimulation, *s,* measured in dollars. Thus, s is added to the physician's exogenously given base wage, w.

The two conventional arguments of the physician's utility function are income, y, and hours worked, h, where $y = h(w + s)$. Stimulation enters the utility function negatively, since activities such as boosting prices and skimping on care are unappealing. Reference income, r, is the final component of utility. Thus, we have the physician's utility level, u, as $u = U(y, h, s, r)$. First derivatives are positive with respect to y, and negative with respect to h and r. All second derivatives are negative. Reference income, r, affects the marginal utility of income. When $r > y$ we say there is an earnings shortfall, and additional income is valued much more highly than when $y \geq r$. This break in utility at a reference point is called loss aversion.[25]

3.1. Loss aversion conjecture

Differences in the marginal utility of income lead physicians with $y < r$ to make more significant attempts to raise their future incomes than do physicians with $y \geq r$. These efforts may involve working more hours and/or engaging in more stimulation.

In sum, when y is below r, we expect to see increased s, which will in turn increase future income.

4. Hypotheses, empirical models, and econometric issues

In earlier work that studied males only, we found strong empirical support for reference income behavior (2003). In the YPS data set examined, actual earnings for period t are reported in period $t + 1$, as is the reference income for period $t + 1$. Using this information, we estimated the effects of reference incomes on the future earnings of self-employed physicians, as well as on hourly wages

[24] In fact, a much earlier literature considered related concepts (interpersonal comparisons of income and consumption, as well as intrapersonal longitudinal comparisons of income and consumption over time) to be important predictors of behavior (see for example Brady and Friedman, 1947; Duesenberry, 1949; Johnson, 1952).

[25] See (2003) for the complete model of reference-income behavior, which builds on Prospect Theory, developed by Kahneman and Tversky (1979).

and hours worked. We estimated two equations, one for the below-reference-income cohort and one for the at-or-above-reference income cohort, each of the form:

$$\ln \text{INCOME90} = a_1 + a_2 \ln \text{INCOME86} + a_3 \ln \text{REFINC87} + \mathbf{aX} + e, \quad (1)$$

where ln indicates the natural logarithm operator, and **X** is a vector of physician-specialty dummies, practice characteristics, and demographic characteristics.

We hypothesized that reference incomes predict earnings growth for male physicians. Specifically, we conjectured that a_3 would be positive and economically and statistically significant for the below-reference-income cohort, and that a_3 for the above-reference-income cohort would be significantly smaller than it was for the below-reference cohort. Both of these hypotheses were confirmed.

4.1. Empirical hypotheses on gender differences

Our principal questions are whether female and male doctors behave differently, and if so, whether such differences can be traced to reference income behaviors. Since reference incomes exert a much more powerful effect in the below-reference-income cohort, which comprises two-thirds of the population, we conduct the rest of our analysis using data for doctors in this category. Appendix B provides parallel analyses for the entire sample.

Differences in preferences between men and women could be of two types. First, women might set lower reference incomes. Second, they might respond less vigorously to shortfalls from their reference incomes. To address these possibilities, we test three major hypotheses:

Hypothesis 1. When actual incomes fall short of reference incomes, males respond strongly; females do not.

Hypothesis 2. Males set higher reference incomes than females.

Hypothesis 3. Differences in reference income behavior (i.e., setting and responding to reference incomes) account for gender differences in earnings growth and absolute earnings.

Two alternative explanations for our observed patterns are considered briefly:

Hypothesis A. Productivity, as measured by indicators of qualifications and human capital, does not explain gender differences in young physicians' earnings growth or absolute earnings.

Hypothesis B. Prejudice, as indicated by reports of discrimination, does not explain gender differences in young physicians' earnings growth or absolute earnings.

To investigate whether male and female physicians respond differently to reference incomes, we estimated the regression:

$$\ln \text{INCOME90} = b_1 + b_2 \ln \text{INCOME86} + b_3 \ln \text{REFINC87} + b_4 \text{FEMALE} + b_5 \ln \text{REFINC} \times \text{FEMALE} + \mathbf{bX} + \eta, \quad (2)$$

where FEMALE takes the value 1 if female and 0 if male. Note that we include a term interacting reference income and physician gender. The coefficient on reference income alone indicates the response of male doctors. The female doctors' response is the sum of the coefficient on reference income and the coefficient on the interaction term. Eq. (2) thus tests the hypotheses that reference incomes affect men's incomes and that they affect women's incomes less than men's. That is, $b_3 > 0$ and $b_5 < 0$. To test whether reference income behavior affects the gender earnings gap, we

compare the gender earnings gap after accounting for reference income effects to the differential without these effects.

As a subsidiary investigation, we also examine how physicians change their behaviors in pursuit of reference points (whether they increase hours worked, earnings per hour, or both). The equations we estimate to study changes in hours and wages are similar to those specified in (1) through (2), but with income in 1990 replaced with hours and wages, respectively. In addition, we add hours in 1986 to the right hand side for the hours' equation. That way the reference income effects estimated by these equations will capture effects on *changes* in hours; we do the same for the wages equation.

4.2. Econometric issues

While our estimation methods are straightforward, a number of econometric issues warrant attention. These include potential biases resulting from measurement errors in the reference income variable, omitted variables, and self-selection effects.[26] As we discuss in (2003), we believe that these issues do not pose serious problems. First, there is little reason to suspect that measurement error in the reference income variable is particularly severe or that it differs systematically between male and female physicians. While omitted variables bias is always a potential problem in empirical research, our concerns about this issue are diminished since we control for a number of sociodemographic, human capital, and geographic factors that typically appear in earnings equations. However, there may be gender-related omitted variables. That is, males and females may differ in important ways that are difficult to measure. For example, female physicians in our sample may be less productive than male physicians. Below, we address this issue to the extent the data permit by interacting ln REFINC87 with productivity-related measures in addition to FEMALE.[27] This allows us to determine whether gender differences in reference income effects persist even after we control for productivity differences that might affect physicians' responsiveness to reference incomes.

Finally, selection problems could arise from utilizing only the below reference income sample. Who falls into the above and below groups depends on characteristics and choices, surely not random assignment. To assess whether selection problems might have biased our below-sample results, we conducted a completely parallel analysis employing the entire sample, doctors both above and below their reference incomes. That analysis reaches the same conclusion: females do not respond to reference incomes; males respond strongly, particularly when they are below their reference incomes.[28]

Despite the statistical significance of our results, our restricted sample size, particularly for female physicians (398 males and 66 females below RI), no doubt prevented us from finding other relationships of interest. A deeper concern is that there could be simultaneity between reference

[26] In principle, each of these issues may be addressed by instrumental variables (IV) estimation. However, in the absence of a theory to help specify an IV model and a data set that includes plausible instruments, the application of IV methods would likely lead to specifications errors. As a number of studies in the econometrics literature indicate (Nelson and Startz, 1990; Bound et al., 1995; Staiger and Stock, 1997), IV estimation can produce substantial biases when, as in our study, available instruments are weak.

[27] The productivity-related measures we consider are binary variables indicating whether the physician is in a surgical or nonsurgical specialty and board certification status. On the average, surgical specialists have significantly longer training periods and earn higher incomes than do nonsurgical specialties. Board certification status is an indicator that a physician is proficient in his or her specialty.

[28] These results are available on the internet at: www.elsevier.com/locate/econbase.

and actual incomes. For example, our reference income measure might merely serve as a proxy for predicted income, quite possibly some form of permanent income measure. If so, we would expect there to be little association between reference and actual income once predicted income had been controlled for. Unfortunately, the other variables in our sample did not provide appropriate instruments for estimating reference income. Hence an instrumental variables approach was not feasible. In (2003), we address the simultaneity issue by first predicting income in 1990 without using reference income, and then regressing actual 1990 income on reference income in 1987 and predicted income in 1990. Those results showed that reference income still had an important effect on income growth after predicted income had been controlled for, suggesting that our reference income measure is substantially more than an income forecast.

5. Results

5.1. Test of Hypothesis 1—males respond strongly to earnings shortfalls; females do not

Multivariate estimates of income growth are given in Table 4. Column 3 shows the results for the estimated model specified in Eq. (2) above, while column 1 shows a traditional earnings model that does not include reference income effects. To provide a comparison, column 2 shows results for the reference income model given in Eq. (1), which does not allow for differences in reference income behavior between male and female physicians.[29]

The "traditional model" (column 1) shows that if we take no account of reference income, FEMALE is negatively and significantly related to income growth. Correcting for other variables, but leaving reference income aside, we find that female physicians earn 21 percent less than males in a traditional model that ignores reference income effects. Even after reference income is accounted for, as column 2 shows, FEMALE continues its negative significant status. Although the magnitude is diminished, the column 2 estimates still imply that female physicians earn 17 percent less. These are economically significant gaps.

We investigated whether these differential behaviors might arise because women, even those who are doctors, invest more in their children, less in their careers. We added two interaction terms to the independent variables in Table 4, FEMALE × MARRIED and FEMALE × NUMKIDS, to see whether women were affected differently by family status than men. The interaction terms did not approach significance. Including them left our other estimated coefficients virtually unchanged.[30]

[29] Unlike in our earlier study (2003), the present version of Eq. (1) includes a binary indicator variable for physician gender since we now include female physicians as well as males. Note also that gender is interacted with actual as well as reference income because gender-specific differences in the effects of reference income on subsequent income may affect gender-specific differences in the relationship between actual 1986 and actual 1990 income. In particular, if reference income has a more powerful effect on subsequent income for males, we might expect actual income to be less strongly related to subsequent income. As an extreme example to illustrate this possibility, suppose that reference income effects were very strong, such that male physicians always hit their reference incomes over time. Then we would expect variations in subsequent income among male physicians to be entirely related to variations in reference income; realized 1986 income would have no influence on subsequent 1990 income.

[30] The MARRIED interaction term was negative; the NUMKIDS interaction term was positive. Results available from the authors. Fuchs (1988), best known for the "invest in kids" explanation and looking across occupations and skill levels, also noted that women tend to get segregated into certain occupations. Hence, it is not surprising that this does not explain behavior by females in a male-dominated, highly skilled profession.

Table 4
Gender effects on 1990 income, including reference income effects ($n = 464$; t-statistics in parentheses)

	Traditional model (no reference income effects) adj. $R^2 = 0.44$	Reference income effects adj. $R^2 = 0.50$	Reference income effects, differentiated by gender adj. $R^2 = 0.50$
INTERCEPT	2.74*** (13.81)	1.67*** (6.98)	1.48*** (5.83)
ln INCOME86	0.47*** (9.70)	0.14** (2.22)	0.13 (1.94)
ln REFINC87	–	0.53*** (7.21)	0.58*** (7.60)
FEMALE	−0.21*** (3.72)	−0.17*** (3.09)	1.03 (1.89)
ln INCOME86 × FEMALE	–	–	0.19 (0.87)
ln REFINC87 × FEMALE	–	–	−0.45[a] (1.94)
MARRIED	0.10 (1.68)	0.08 (1.45)	0.08 (1.40)
NUMKIDS	0.01 (0.56)	0.02 (0.97)	0.02 (0.96)
PRACSIZE	−0.001 (1.03)	−0.0003 (0.47)	−0.0003 (0.43)
PRACYR	−0.06*** (3.64)	−0.05*** (3.10)	−0.04*** (3.04)
GENIM	0.16** (2.43)	0.13** (2.09)	0.13** (2.04)
PED	0.19*** (2.72)	0.20*** (2.98)	0.18*** (2.73)
SPECIM	0.23*** (3.11)	0.14 (1.96)	0.13 (1.81)
GENSURG	0.46*** (4.70)	0.28*** (2.97)	0.26*** (2.73)
SPECSURG	0.49*** (7.17)	0.37*** (5.56)	0.36*** (5.39)
PSYCH	0.07 (0.61)	0.10 (0.95)	0.09 (0.88)
OBGYN	0.62*** (7.53)	0.54*** (6.80)	0.54*** (6.87)
BLACK	−0.07 (1.07)	−0.06 (1.02)	−0.06 (1.00)
BDCERT	0.09** (2.00)	0.10** (2.24)	0.10** (2.29)
FMG	0.003 (0.05)	−0.02 (0.32)	−0.02 (0.25)
RURAL	−0.04 (0.40)	0.001 (0.01)	−0.01 (0.12)
NONSMSA	−0.04 (0.69)	−0.01 (0.16)	−0.01 (0.24)
SMALSMSA	−0.04 (0.76)	−0.02 (0.41)	−0.03 (0.68)
BIGSMSA	−0.04 (0.71)	−0.05 (0.85)	−0.06 (1.09)
MIDWEST	−0.04 (0.67)	−0.02 (0.29)	−0.02 (0.36)
SOUTH	−0.02 (0.35)	0.002 (0.03)	0.01 (0.12)
WEST	−0.08 (1.33)	−0.07 (1.30)	−0.07 (1.27)

[a] An F-test for joint significance for ln REFINC87 + ln REFINC87 × FEMALE: not significant ($p = 0.38$).
** Statistically significant at the 5 percent level, two-tailed test.
*** Statistically significant at the 1 percent level, two-tailed test.

Let us return to Hypothesis 1, which was that males respond more strongly than females to earnings shortfalls. The response of females is given by the sum of direct and indirect effects, namely, the sum of the coefficients on ln REFINC87 and ln REFINC87 × FEMALE. The results in column 3 provide the evidence: they reveal two striking patterns. First, we find that reference incomes are directly related to earnings growth for males but not for females.[31] Second, the direct effect of FEMALE on earnings growth is no longer statistically significant. Calculations taking account of total direct and indirect effects (evaluated at sample mean values for reference and actual incomes) show that females earn on average 22 percent less than males. This is consistent with the results in column 1 for the traditional model. What the column 3 results further suggest, however, is that the gender earnings differential may be traced to gender-specific differences in the effects of reference income on subsequent incomes. Collectively, these findings suggest that

[31] The F-test for the joint significance of ln REFINC87 and ln REFINC87 × FEMALE (i.e., the coefficient 0.58–0.45) gives the estimated effect of reference incomes for females. It is far from statistically significant ($p = 0.38$). This finding is not surprising since the sum of the estimated coefficients on these two variables is close to zero.

Table 5
Gender effects on wages and hours worked

	No reference income effects	Reference income	Reference income, differentiated by gender
Wage effects			
ln REFINC87	–	0.45*** (5.16)	0.49***[a] (5.33)
FEMALE	−0.13 (1.92)	−0.10 (1.63)	1.09 (1.63)
ln REFINC87 × FEMALE	–	–	−0.25[a] (0.92)
Hours worked			
ln REFINC87	–	0.03 (0.65)	0.05[b] (1.06)
FEMALE	−0.07 (1.96)	−0.07 (1.91)	0.08 (0.22)
ln REFINC87 × FEMALE	–	–	−0.26[b] (1.72)

[a] An F-test for joint significance for wage equation: ln REFINC87 + ln REFINC87 × FEMALE: not significant ($p = 0.35$).
[b] An F-test for joint significance for hours worked equation: ln REFINC87 + ln REFINC87 × FEMALE: not significant ($p = 0.16$).

male physicians engage in reference income behavior but female physicians do not, certainly not strongly. Hypothesis 1 is confirmed.

5.1.1. Gender effects on hourly wages and hours worked

How do males raise their actual incomes in response to income shortfalls? Do they raise hourly earnings, or work more hours, or both? Presumably, given our findings, females do not respond significantly in either of these areas. To address these issues, we examined how reference income influenced wages and hours worked for men and for women.

Basically, we employed the multivariate models described in Eq. (2) above, with wages or hours worked replacing income as the dependent variable. These results are given in Table 5.[32] Reference incomes are unrelated to hours worked for both males and females. By contrast, gender-specific differences in reference income behavior significantly affect hourly wages. While ln REFINC87 has a positive and highly significant wage effect for males, an F-test for ln REFINC87 and its interaction with FEMALE cannot reject the hypothesis that these two variables are jointly insignificant, that is, that reference incomes are unrelated to wage growth for females.

5.2. Test of Hypothesis 2—males set higher reference incomes

As Table 2 indicates, the mean reference incomes of males are significantly higher than those of females in both the full and below-reference-income samples; these differences are significant at the 1 percent level.

Correcting for background variables, do these differences persist? In results not shown, we regressed r on our independent variables, but we did not include actual income. The coefficient on FEMALE showed that women doctors set significantly lower reference incomes, holding other factors fixed. Hypothesis 2 is confirmed.

[32] The results in Table 5 are based on multivariate models that include all of the right-hand side variables in the income equations (see Table 4). Moreover, for the hours worked equations, we included hours worked during the previous period as an additional explanatory variable. For the wage equations, we used wages from the previous period. The reason for including lagged hours worked and wages is to isolate the effects of reference income on changes in hours worked and wages, so as to be consistent with our main analysis focusing on annual income changes.

Table 6
Disentangling gender differentials in reference income behavior

Assumed reference income level	Assuming female reference income behavior	Assuming male reference income behavior
Predicted female incomes[a]		
Female reference income	$111000***	$132000***
Male reference income	$114000***	$153000
Predicted male incomes[b]		
Male reference income	–	$157000

[a] These estimates show how differences in reference income levels and reference income behavior would affect the earnings of female physicians. They use the model coefficients reported in column 3 of Table 4. The female sample values generate predicted values of income for each female in the sample, setting reference income to either the mean value for females or males, and using behavioral assumptions for females or males. For example, the predicted income assuming female reference income and female reference income behavior ($111,000) was generated by using all of the coefficients in column 3 of Table 4, including the coefficients on FEMALE and its interactions with reference and actual income. Reference income was set to the mean value for females, and values of other variables were specific to each female in the sample. The figure of $111,000 is thus the mean income of female physicians in the sample, if they each had reference incomes equal to the sample mean for females and engaged in female reference income behavior. The female behavior/male reference income value ($114,000) was calculated in a similar manner, but replaced average female reference income with average male reference income. The female reference income/male reference income behavior calculation ($132,000) used the mean value of female reference income, and omitted coefficients on FEMALE and its interactions with reference and actual income. The male reference income/male reference income behavior model ($153,000) used the mean value of male reference income, and omitted coefficients on FEMALE and its interactions with reference and actual income.

[b] This estimate used the mean value of male reference income, and omitted coefficients on FEMALE and its interactions with reference and actual income, generating predicted values of income for each male in the sample. The $157,000 figure is the mean value of predicted income for males in the sample.

*** This value is significantly lower than the predicted income of males ($157,000) at the 1 percent level.

5.3. Test of Hypothesis 3—reference income behavior explains gender differences in absolute earnings

Reference income effects may differ by gender for two reasons. First, females set lower reference points, as Table 2 suggests and the test of Hypothesis 2 confirms. Second, females respond at most modestly to reference incomes; see column 3 in Table 4. To disentangle these effects, we use the regression results from column 3. We wish to determine what the earnings of female physicians would be under different assumptions regarding reference income levels and reference income behaviors. For each doctor, we use the model coefficients reported in column 3 of Table 4 and the actual values of the other independent variables, including 1986 actual income. We then estimate the amount each female doctor would earn if she had the mean reference income of females or males, first responding according to the coefficient on reference income for females, and then according to the one for males.[33] Thus, for each female doctor we have four estimates in a 2×2 framework. Finally, for purposes of comparison, we calculated predicted incomes for males using mean values for males in the sample, and applying male reference income levels and male reference income behavior.

These results, summarized in Table 6, indicate that the gender-specific difference in reference income behavior is the main reason why females earn less than males. Using female

[33] See the legend to Table 6 for a fuller description of the methodology used to obtain these estimates.

reference income levels, hypothetical women who engage in male reference income behavior increase their average earnings from $111,000 to $132,000.[34] Using the higher male reference incomes, the increase is even greater, rising from $114,000 to $153,000. This $153,000 figure is close to the $157,000 computed amount that males would earn using male sample values and reference income behavior. To test whether males and females had different earnings apart from reference income behavior, we looked at the distributions of estimated incomes for male and female doctors, assuming male reference income and responsiveness for both groups. Given the sample sizes and their standard deviations ($31,900 and $36,400), the difference in means is far from significant ($p = 0.47$). Hypothesis 3 is confirmed. Differences between males and females in setting and responding to reference incomes explain virtually all of the gender differences in the absolute earnings of young physicians.[35]

Though Hypothesis 3 appears to explain male-female earnings patterns, we consider two common alternative explanations, productivity differentials and discrimination, to the extent our data permit.

5.4. Test of Hypothesis A—productivity measures do not explain differentials in earnings growth or absolute earnings

If productivity measures significantly affect reference income behavior, there could be an interaction effect. We lack the data for a full analysis, but we did look at whether specialty mix and board certification were correlated with reference income behavior. There are disproportionately few females in the surgical specialties, which involve lengthy training periods (Bowman and Gross, 1986). If reference income behavior differs between surgical and non-surgical specialties, perhaps because consumers have poor information about the appropriateness of surgical procedures or because surgeons have a broader portfolio of procedures to choose from in striving to attain their reference points, then the gender-specific difference in reference income behavior may derive at least partly from gender-specific differences in specialty mix. To test this hypothesis, we interacted ln REFINC87 with a binary variable equal to 1 if the physician is in a surgical specialty and 0 otherwise. Board certification status is a second indicator of human capital that we investigate to understand gender differences in earnings. We interacted board certification status with ln REFINC87 and added this term to our income model. Other variables included in this model are the same as in Table 4.

Neither of these interaction terms, both of them being human capital measures, significantly affects the gender earnings gap.[36] In our test of Hypothesis 3, we found no significant difference in male and female incomes, assuming male reference income behavior for both. In addition, two important measures of human capital proved not to affect gender differences in reference income behavior. Together, these findings support Hypothesis A.

[34] In the absence of male reference income behavior, however, higher reference incomes alone do little to increase the earnings of females, raising income from $111,000 to $114,000.

[35] These patterns persist in models that pool the above and below reference income physicians and include interaction terms to allow reference income effects to differ by gender and reference income status (see Appendix B).

[36] These results are omitted in the interest of brevity. However, they are available at the following website: www.elsevier.com/locate/econbase.

5.5. Test of Hypothesis B—prejudice does not explain differentials in earnings growth or absolute earnings

Discrimination is often posited to explain gender earnings differentials. Employers, patients, or even peers might all discriminate. Discrimination by peers would be unusually significant in the medical profession, given the importance of referrals, and if it happens early in a physician's career, say during residency, it could limit professional opportunities later.

Discrimination is often measured merely as a residual: after all other factors are taken into account, is the coefficient on FEMALE still negative and significant? Here the answer, provided in Table 4, is no, but we should be cautious before taking reassurance from this result. Residual tests of discrimination are indirect, and whatever the results, we would prefer a direct test. Fortunately, the YPS provides a more direct test for the effects of discrimination. The 1991 survey asked doctors whether, during the course of their training and medical careers, their opportunities were ever limited because of gender. If people who answer yes to this question had lower income or income growth than people who answer no, that would indicate that prejudice helped explain gender differences in earnings.[37] By contrast, if there were no difference, then job- and training-related prejudice, though still possibly present, did not contribute to gender earnings differentials. This study is limited by the reliability of self-reports.

Thirty-four of the 66 female doctors reported suffering discrimination.[38] Did these doctors, denoted as "reporters" below, fare worse than those who did not, labeled "non-reporters"? In raw terms, reporters and non-reporters do not differ significantly in incomes in either 1986 or 1990 or in reference incomes. It is possible, however, that differences would emerge if we controlled for differences in background variables. To examine this possibility, we recomputed our base equation, column 3 in Table 4, adding two dummy variables: FEMNODISC takes a value of 1 for females who do not report, otherwise 0; FEMDISC takes a 1 for reporters, otherwise 0.

We estimated a model that enters FEMNODISC and FEMDISC separately, then interacts these two variables with reference and actual income.[39] We find that the impact of reference income does not vary according to whether females believed that they had been discriminated against or not. Indeed, these interaction terms are very similar in magnitude, and we cannot reject the hypothesis that reference income is unrelated to income growth for each group of females.

[37] What if reported discrimination is a poor indicator of actual discrimination? For physicians at least, Hampton and Heywood (1993) find that perceived gender discrimination correlates strongly with econometric estimates of wage discrimination. However, Kuhn (1987) and Barbezat and Hughes (1990) have found that perceived discrimination tends to be lowest in cases where measured salary discrimination is greatest. Although this finding was statistically insignificant in the Kuhn analysis and insignificant in most cases in the study by Barbezat and Hughes, the negative correlation was consistent across these studies. The precise reasons for this relationship are unclear. Barbezat and Hughes suggest that employers may tend to engage in gender discrimination when it is unlikely that they will be detected. Kuhn suggests that female workers might use other criteria in addition to salary differentials when determining whether they have been discriminated against.

[38] Such reports were virtually absent among male respondents; less than 2 percent report having incurred any gender discrimination. A greater proportion of males (54 percent) than females (35 percent) are self-employed. If self-employment is advantageous, this disproportionate share of self-employed males may also indicate discrimination against females. We thank a referee for calling this point to our attention.

[39] Again in the interest of brevity, these regression results are omitted but are available at: www.elsevier.com/locate/econbase.

Moreover, we could not reject the hypotheses that FEMNODISC and FEMDISC did not differ significantly.[40]

However, we have looked only at growth rates of income. One might argue that our finding that income growth rates do not differ among females according to whether they perceived that they were discriminated against does not provide a complete test of discrimination because females subject to discrimination early on may have had lower prior levels of income. If so, this would be reflected in 1986 incomes. We examined these issues, comparing actual incomes for females according to whether or not they perceived any gender discrimination. We regressed y on our independent variables (excluding actual income), adding binary variables that indicated whether females perceived gender discrimination or not.

We find that reporters have income levels identical to males.[41] Non-reporters do significantly worse than males. This pattern works against a discrimination finding. However, the sample is relatively small, and an F-test could not reject the hypothesis that earnings do not differ significantly between reporters and nonreporters ($p = 0.08$).[42]

Reporters and non-reporters have similar earning experiences. Discrimination, at least as perceived by the female physicians themselves, does not lead to lower incomes or growth in incomes. Neither does it explain the lack of a relationship between ln REFINC87 and income growth for females. This evidence is consistent with Hypothesis B. Nonetheless, the results on discrimination should be interpreted with caution since they are based on self-reports.

Quite apart from our hypotheses and statistical tests, it is noteworthy that half the female doctors reported suffering discrimination in their careers. Whatever the economic consequences associated with these reports, and whatever the stimuli that produced them, this finding is disturbing.

6. Conclusion

Why do male doctors earn more than their female counterparts? Why do their earnings grow faster? We find that gender differences in preferences account for the entire differentials in income and income growth rate.[43] Our preference analysis revolves around two critical concepts from prospect theory, the existence of a reference point for measuring gains and losses, and loss aversion. We posit that physicians' reference incomes (what they report to be adequate incomes given their career stages) serve as a reference point, and loss aversion related to shortfalls strongly encourages them to push toward that point. However, male and female doctors differ significantly in reference income behavior. Males set higher reference points, and when they perceive their incomes to be

[40] An F-test for the difference between FEMNODISC and interaction terms involving FEMNODISC, and FEMDISC and interaction terms involving FEMDISC was not statistically significant ($p = 0.15$). This is consistent with Antecol and Kuhn (2000),who study specific reports of discrimination, finding that young female job seekers are three times as likely to report that job finding is impeded by gender. They find these reports unrelated to any objective measure of discrimination.

[41] Reference incomes may also differ depending on whether females perceived that they were discriminated against. Thus we also predicted 1987 reference incomes. With actual incomes excluded as independent variables, the coefficients for FEMDISC and FEMNODISC are negative, and for FEMNODISC, statistically significant at the 1 percent level. If anything, reporters have higher reference incomes, although an F-test could not reject the hypothesis that the coefficients on FEMDISC and FEMNODISC were not significantly different from each other ($p = 0.17$).

[42] For the full set of these results, see: www.elsevier.com/locate/econbase.

[43] Our available measures of discrimination and productivity are imperfect, however, since discriminatory experience was determined by a self-report. Quite possibly, more refined measures or a larger data set could reveal that these factors contribute to gender disparities in actual and reference incomes.

inadequate, they change behavior to close the gap. In particular, when males have an income shortfall, their stimulation efforts enable them to earn more per hour, whether by spending less time per patient or focusing on more lucrative procedures. Other studies have associated these behaviors with male physicians.

Neoclassical economists are understandably reluctant to ascribe differences in behavior to differences in preferences, because preferences are rarely observable, and such an "explanation" can be applied to almost all observed phenomena. Assuming full rationality leads to much tighter predictions of behavior. By contrast, the new behavioral economics, exemplified by the work of Kahneman and Tversky (1979) and Thaler (1991), stresses a role for preferences and allows for inconsistencies within individuals, and for behaviors far from what rational theory would predict. Neither approach attends much to systematic differences in preferences across classes of individuals an important possibility raised by this analysis.

Clearly, a balance must be struck between the two views. One should amend neoclassical models with caution, only when the empirical evidence for doing so is compelling and likely to provide new insights. However, failure to consider empirical evidence creates the risk of living with elegant but unsupported theories. Our study seeks to augment standard theory, not to contradict it. We believe that existing evidence on labor supply behavior and preferences over job characteristics, as well as evidence on medical practice patterns, point to substantive gender-specific differences in physicians' behavior. This behavioral hypothesis, in conjunction with earnings equations that are otherwise quite standard, shows that female physicians are less willing than males to bend their behaviors to push toward reference incomes. Such behavioral differences, when coupled with the higher reference incomes set by males, are sufficient to explain the gender gap in physicians' incomes. Thus, efforts to reduce gender discrimination and promote human capital among female physicians (for example, to increase the number of female surgical specialists) may not noticeably reduce gender differences in earnings.[44]

Our findings suggest that policymakers need to know more about the consequences of reference income behavior for patients. Does the absence of reference income behavior among female physicians indicate that they are better agents for their patients and less likely to induce medically marginal demand? If so, then it would surely not be in society's interest to lure female physicians to mimic their male counterparts' emphasis on reference incomes. Or perhaps male physicians engage in efforts to achieve reference incomes without inducing demand, and female physicians' reluctance to do so may reflect socialization that restricts their autonomy and confidence in their own decisions. If so, it may be desirable for female physicians to behave more like males with respect to reference income behavior.

The underpinnings for these policy questions are complex and unclear. What is clear, however, is that lessons from psychology and medical care research, when blended with economic theory and empirical assessment, can help explain gender-specific differences in the economic behavior

[44] In recent years, the government has employed RBRVS (resource-based relative value scales) to redistribute its Medicare payments for various types of physician services. The philosophy has been to base payments on the skill and effort a treatment requires rather than some customary figure that might reflect such factors as market power. Some might mistakenly conclude that there would be little cost in extending this system to raise female physicians' income in the spirit of comparable worth. However, unless resources are the basis for payment, inefficiencies will intrude. Moreover, female physicians might ultimately not receive most of the benefits. For example, Orazem and Mattila (1990, p. 149), studying comparable worth plans in Iowa, found that potential pay gains of 8.8 percent for female employees dwindled to 1.4 percent once the system was fully implemented and appeals resolved.

of physicians. That understanding is critical if we are to determine what policy measures, if any, are warranted.

Acknowledgements

We thank Kenneth Arrow, Miriam Avins, Francine Blau, David Bradford, George Borjas, Jim Burgess, Carol Chetkovich, Suzanne Cooper, David Cutler, Sherman Folland, Jane Fountain, Alain Enthoven, Victor Fuchs, Karen Goldsteen, Ray Goldsteen, Tom Getzen, Jon Gruber, George Jakubson, Rob Jensen, Ted Joyce, Daniel Kahneman, Tom Kane, John Kovach, David Laibson, Dorothy Lane, Albert Ma, Jane Mansbridge, Lucille Schmidt, Miron Stano, Warren Sanderson, James Stock, Paul Todgham, Julie Wilson, Paul Wilson, and participants in seminars at Baruch College, Florida International University, Stanford, Stony Brook, Berkeley, the Medical University of South Carolina, Oakland University, the Chinese Academy of Preventive Medicine, Beijing, and the Northeast Regional Health Economics Research Symposium for helpful comments and suggestions. We thank the Women and Public Policy Program, Harvard University, for research support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2005.06.007.

References

Antecol, H., Kuhn, P., 2000. Gender as an impediment to labor market success. Journal of Labor Economics 18, 702–738.

Barbezat, D., Hughes, J., 1990. Sex discrimination in labor markets: the role of statistical evidence (comment). American Economic Review 80, 277–286.

Bashaw, F., Heywood, J., 2001. The gender earnings gap for US physicians: has equality been achieved? Labour 15, 371–391.

Bickell, N.J., Earp, J., Garrett, J., Evans, A., 1994. Gynecologists' sex, clinical beliefs, and hysterectomy rates. American Journal of Public Health 84, 1649–1652.

Blinder, A., 1973. Wage discrimination: reduced form and structural estimates. Journal of Human Resources 9, 436–455.

Bobula, J., 1980. Income differences between male and female physicians. In: Glandon, G., Shapiro, R. (Eds.), Profile of Medical Practice. American Medical Association, Chicago, IL, pp. 123–133.

Bound, J., Jaeger, D., Baker, R., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association 90, 443–450.

Bowlus, A., 1997. A search interpretation of male–female wage differentials. Journal of Labor Economics 15, 625–657.

Bowman, M., Gross, M., 1986. Overview of research on women in medicine: issues for policymakers. Public Health Reports 101, 513–520.

Bowman, D., Minehart, D., Rabin, M., 1999. Loss aversion in a consumption-savings model. Journal of Economic Behavior and Organization 38, 155–178.

Brady, D., Friedman, R., 1947. Savings and the income distribution. Studies in Income and Wealth, Conference on Research in Income and Wealth, vol. 10. National Bureau of Economic Research, New York, pp. 247–265.

Brannon, L., 1996. Gender: Psychological Perspectives. Allyn and Bacon, Needham Heights, MA.

Britt, H., Bhasale, A., Miles, D., Meza, A., Sayer, G., Angelis, M., 1996. The sex of the general practitioner: a comparison of characteristics, patients, and medical conditions managed. Medical Care 34, 403–415.

Camerer, C., 2003. Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton, NJ.

Culler, S., Ohsfeldt, R., 1986. Differences in income between male and female physicians. Journal of Health Economics 5, 335–346.

Daymont, T., Andrisani, P., 1984. Job preferences, college major, and the gender gap in earnings. Journal of Human Resources 19, 412–413.

Domenighetti, G., Luraschi, P., Marazzi, A., 1985. Hysterectomy and the sex of the gynecologist (letter). New England Journal of Medicine 313, 1482.

Duesenberry, J., 1949. Income, Savings, and the Theory of Consumer Behavior. Harvard University Press, Cambridge, MA.

Dunn, L., 1996. Loss aversion and adaptation in the labor market: empirical indifference functions and labor supply. Review of Economics and Statistics 78, 441–450.

Frank, R., 1985. Choosing the Right Pond: Human Behavior and the Quest for Status. Oxford University Press, Oxford, UK.

Fuchs, V., 1988. Women's Quest for Economic Equality. Harvard University Press, Cambridge, MA.

Gilligan, C., 1982. In a Different Voice: Psychological Theory and Women's Development. Harvard University Press, Cambridge, MA.

Hampton, M., Heywood, J., 1993. Do workers accurately perceive gender wage discrimination? Industrial and Labor Relations Review 47, 36–49.

Harvard Health Letter, 2001. The doctor–patient relationship, doctors from Venus and Mars: how they differ 26, 4. Consumer Health Publishing Group, Stamford, CT.

Johnson, D.K., 1988. Adolescents' solutions to dilemmas in fables: two moral orientations – two problem solving strategies. In: Gilligan, C., Ward, J., Taylor, J., Bardage, B. (Eds.), Mapping the Moral Domain: A Contribution of Women's Thinking to Psychological Theory and Education. Harvard University Press, Cambridge, MA, pp. 49–71.

Johnson, H., 1952. The effect of income-redistribution on aggregate consumption with interdependence of consumers' preferences. Economica 19, 131–147.

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–291.

Kahneman, D., Tversky, A., 1984. Choices, values, and frames. American Psychologist 39, 341–350.

Kehrer, B., 1976. Factors affecting the incomes of men and women physicians: an exploratory analysis. Journal of Human Resources 11, 526–545.

Kennedy, P., 1992. A Guide to Econometrics, 3rd ed. MIT Press, Cambridge, MA.

Killingsworth, M., 1983. Labor Supply. Cambridge University Press, Cambridge, UK.

Kuhn, P., 1987. Sex discrimination in labor markets: the role of statistical evidence. American Economic Review 77, 567–583.

Langwell, K., 1982. Factors affecting the incomes of men and women physicians: further explorations. Journal of Human Resources 17, 261–274.

Lurie, N., Slater, J., McGovern, P., Ekstrom, J., Quam, L., Margolis, K., 1993. Preventive care for women: does the sex of the physician matter? New England Journal of Medicine 329, 478–482.

Lyons, N., 1988. Two perspectives: on self, relationships, and morality. In: Gilligan, C., Ward, J., Taylor, J., Bardage, B. (Eds.), Mapping the Moral Domain: A Contribution of Women's Thinking to Psychological Theory and Education. Harvard University Press, Cambridge, MA, pp. 21–48.

Martin, S., Arnold, R., Parker, R., 1988. Physician gender and medical socialization. Journal of Health and Social Behavior 29, 333–343.

Mitchell, J., 1982. Practice patterns and earnings of women physicians. Discussion Paper, Boston University.

Mitler, L., Rizzo, J., Horwitz, S., 2000. Physician gender and caesarian sections. Journal of Clinical Epidemiology 53, 1030–1035.

Nelson, C., Startz, R., 1990. The distribution of the instrumental variables estimator and its *t*-ratio when the instrument is a poor one. Journal of Business 63, S125–S140.

Oaxaca, R., 1973. Male–female wage differentials in urban labor markets. International Economic Review 14, 693–709.

Orazem, P., Mattila, J., 1990. The implementation process of comparable worth: winners and losers. Journal of Political Economy 98, 134–152.

Rizzo, J., Blumenthal, D., 1994. Physician labor supply: do income effects matter? Journal of Health Economics 13, 433–453.

Rizzo, J., Zeckhauser, R., 2003. Reference incomes, loss aversion, and physician behavior. Review of Economics and Statistics 85, 909–922.

Staiger, D., Stock, J., 1997. Instrumental variables estimation with weak instruments. Econometrica 65, 557–586.

Thaler, R., 1980. Toward a positive theory of consumer choice. Journal of Economic Behavior and Organization 1, 39–60.
Thaler, R., 1991. Quasi-rational Economics. Russell Sage Foundation, New York.
Wertz, D., 1993. Provider biases and choice: the role of gender. In: Evans, M. (Ed.), Clinical Obstetrics and Gynecology. Year Book Medical Publishers, Chicago, IL, pp. 521–531.
Wood, R., Corcoran, M., Courant, P., 1993. Pay differences among the highly paid: the male–female earnings gap in lawyers' salaries. Journal of Labor Economics 11, 417–441.