Chen JX, Chang EH, Deng F, et al. Autonomy in the operating room: a multicenter study of gender disparities during surgical training. *J Grad Med Educ*. 2021;13(5):666–672. http://dx.doi.org/10.4300/JGME-D-21-00217.1

Supplementary Data

SUPPLEMENTARY METHODS

Fixed effects were used to provide conservative estimates of the effects of trainee gender on dependent variables of interest because random effects require assumptions about unobserved heterogeneity being uncorrelated with independent variables of interest, whereas fixed effects require no such assumptions.[1] In addition, regressions were performed with and without fixed effects to show how the estimates of interest change, in case fixed effects throw out useful variation. Standard errors were clustered to account for correlations that arise from repeated observations.[2] Finally, ordinary least squares regressions as opposed to other models like logistic or ordinal regressions were used because they are still robust given the large sample size,[1] while other models require stronger additional assumptions that can lead to bias if violated (e.g., the sign of the interaction in logistic regressions may not correspond to the sign of its effect on the dependent variable).[3]

Chen JX, Chang EH, Deng F, et al. Autonomy in the operating room: a multicenter study of gender disparities during surgical training. *J Grad Med Educ*. 2021;13(5):666–672. http://dx.doi.org/10.4300/JGME-D-21-00217.1

SUPPLEMENTARY TABLES

Table S1. Scale of surgical autonomy. The 4-level Zwisch scale describes attending surgeon and trainee behaviors for >50% of the critical portions of the case (adapted from Chen et al. 2019).[4]

	Level	Attending behaviors	Trainee behaviors
1	**Show and Tell**	Performs majority of key portions as the surgeon Narrates the case, key concepts, anatomy, skills	Opens and closes First assists, observes
2	**Active Help**	Leads actively for >50% of critical portion of case Identifies key anatomy, optimizes the surgical field Coaches technical skills and next steps	Actively assists Practices component technical skills
3	**Passive Help**	Follows trainee's lead for >50% of critical portion Coaches for polish/refinement/safety	Recognizes transition points Can accomplish next steps
4	**Supervision Only**	Gives no unsolicited advice for >50% of critical portion Monitors progress and patient safety	Mimics independence Recovers from most errors

Table S2. Scale of surgical performance. The 5-level performance scale describes trainee performance during surgery (adapted from Chen et al. 2019).[4]

	Level	Performance Descriptor
1	**Unprepared/critical deficiency**	Poorly prepared to performed this procedure and/or included critical performance errors that endangered patient safety or procedure outcomes.
2	**Inexperienced with procedure**	Trainee appears inexperienced in performing this procedure with frequent problems with technique, execution, smoothness, forward planning.
3	**Intermediate performance**	Intermediate stage of development; performance of procedural elements is variable but acceptable for the amount of experience with the procedure; not yet at the level of graduating trainees.
4	**Practice-ready performance**	Trainee is ready to perform this operation safely and independently assuming trainee consistently performs procedure in this manner.
5	**Exceptional performance**	Above the level expected of graduating trainees.

Table S3. Autonomy by training level. Mean autonomy ratings by attending surgeons are listed according to trainee postgraduate year. SD, standard deviation.

Postgraduate Year	Number of Assessments	Mean Autonomy Rating (SD)
1	5,781	1.94 (0.71)
2	5,844	2.32 (0.77)
3	7,881	2.49 (0.78)
4	9,512	2.79 (0.83)
5	8,123	2.95 (0.80)
6	874	2.96 (082)
7	214	3.00 (0.82)
8	706	2.78 (0.78)
9+	686	3.24 (0.73)

Table S4. Autonomy gender gap models. Multivariable regressions model surgical autonomy as rated by attendings based on trainee gender (models 1-4) while accounting for additional fixed effects based on attending (models 2-4), training level (models 3-4), and procedure (model 4). The table shows for each variable the unstandardized regression coefficient B and, in parentheses, the standard errors of B. Significance is denoted with *, **, and *** at $p<0.05$, $p<0.01$ and $p<0.001$ levels, respectively. There was a statistically significant gender gap in the full model (Model 4).

	Model 1	Model 2	Model 3	Model 4
Trainee Gender (1 = Female; 0 = Male)	B= - (0.00891)	-0.0536*** (0.00873)	-0.0146 (0.00798)	-0.0199** (0.00750)
Fixed Effects for Attending	No	Yes	Yes	Yes
Fixed Effects for Training Level	No	No	Yes	Yes
Fixed Effects for Procedure	No	No	No	Yes
Postgraduate Year (PGY) 1			Omitted	Omitted
PGY 2			0.305*** (0.0152)	0.350*** (0.0144)
PGY 3			0.574*** (0.0139)	0.677*** (0.0132)
PGY 4			0.858*** (0.0141)	1.027*** (0.0135)
PGY 5			1.141*** (0.0145)	1.136*** (0.0140)
PGY 6			1.150*** (0.0348)	1.368*** (0.0341)
PGY 7			1.327*** (0.0609)	1.587*** (0.0637)
PGY 8			1.365*** (0.104)	1.509*** (0.103)
PGY 9			1.935*** (0.112)	2.087*** (0.111)
PGY 10			2.227*** (0.546)	3.026*** (0.699)
Observations	39,621	39,621	39,621	39,621
R^2	0.002	0.002	0.157	0.327

Table S5. Autonomy gender gap model with gender–training level interaction. Multivariable regression predicting surgical autonomy as rated by attending was used to evaluate for an interaction between gender and training level, while also accounting for fixed effects for attending, training level, and procedure (analogous to Model 4 in Table S4). The table shows for each variable the unstandardized regression coefficient B and, in parentheses, the standard errors of B. Significance is denoted with *, **, and *** at p<0.05, p<0.01 and p<0.001 levels, respectively. There was a statistically significant interaction between female trainee gender and training level for predicting autonomy.

	Autonomy
Trainee Gender (1 = Female; 0 = Male)	B=0.0266 (0.0269)
Postgraduate Year (PGY)	0.336*** (0.00809)
Female Trainee x PGY	-0.0163* (0.00701)
Fixed Effects for Attending	Yes
Fixed Effects for Training Level	Yes
Fixed Effects for Procedure	Yes
Observations	39,621
R^2	0.288

Table S6. Autonomy gender gap model with gender–case complexity interaction. Multivariable regressions predicting surgical autonomy as rated by attending was used to evaluate for an interaction between gender and case complexity, while also accounting for fixed effects for attending, training level, and procedure (analogous to Model 4 in Table S4). The table shows for each variable the unstandardized regression coefficient B and, in parentheses, the standard errors of B. Significance is denoted with *, **, and *** at p<0.05, p<0.01 and p<0.001 levels, respectively. There was a statistically significant interaction between female trainee gender and case complexity for predicting autonomy.

	Autonomy
Trainee Gender (1 = Female; 0 = Male)	B= -0.0126 (0.00823)
Most Complex Cases (1 if Case Rated as Most Complex; 0 otherwise)	-0.316*** (0.0134)
Female Trainee x Most Complex Cases	-0.0366* (0.0165)
Fixed Effects for Attending	Yes
Fixed Effects for Training Level	Yes
Fixed Effects for Procedure	Yes
Postgraduate Year (PGY) 1	Omitted
PGY 2	0.356*** (0.0142)
PGY 3	0.691*** (0.0129)
PGY 4	1.045*** (0.0133)
PGY 5	1.382*** (0.0138)
PGY 6	1.386*** (0.0336)
PGY 7	1.601*** (0.0627)
PGY 8	1.583*** (0.102)
PGY 9	2.155*** (0.110)
PGY 10	2.977*** (0.688)

Observations	39,621
R^2	0.340

Table S7. Autonomy gender gap subgroup analysis for most complex and less complex cases. Multivariable regressions predicting surgical autonomy as rated by attending was used to separately analyze those cases that were rated "hardest 1/3" (most complex) and those that were rated either "easiest 1/3" or "average 1/3" (less complex), while also accounting for fixed effects for attending, training level, and procedure (analogous to Model 4 in Table S4). The table shows for each variable the unstandardized regression coefficient B and, in parentheses, the standard errors of B. Significance is denoted with *, **, and *** at p<0.05, p<0.01 and p<0.001 levels, respectively. The autonomy gender gap was statistically significant for the most complex cases.

	Less Complex	Most Complex
Trainee Gender (1 = Female; 0 = Male)	B= -0.0107 (0.00841)	-0.0502** (0.0164)
Fixed Effects for Attending	Yes	Yes
Fixed Effects for Training Level	Yes	Yes
Fixed Effects for Procedure	Yes	Yes
Postgraduate Year (PGY) 1	Omitted	Omitted
PGY 2	0.353*** (0.0156)	0.328*** (0.0363)
PGY 3	0.705*** (0.0143)	0.611*** (0.0329)
PGY 4	1.074*** (0.0149)	0.946*** (0.0327)
PGY 5	1.395*** (0.0155)	1.318*** (0.0330)
PGY 6	1.440*** (0.0382)	1.260*** (0.0834)
PGY 7	1.668*** (0.0734)	1.473*** (0.147)
PGY 8	1.638*** (0.132)	1.461*** (0.203)
PGY 9	2.220*** (0.142)	2.026*** (0.216)
PGY 10	2.858*** (0.717)	- -
Observations	30,431	9,190
R^2	0.354	0.291

Table S8. Autonomy gender gap by attending gender for less complex cases. Multivariable regressions predicting surgical autonomy as rated by attending was used to separately analyze female and male attending surgeons in cases rated "easiest 1/3" or "average 1/3" in complexity compared with similar cases (a subgroup analysis of the Less Complex model in Table S7). The table shows for each variable the unstandardized regression coefficient B and, in parentheses, the standard errors of B. Significance is denoted with *, **, and *** at $p<0.05$, $p<0.01$ and $p<0.001$ levels, respectively. The autonomy gender gap was statistically significant for male attendings in less complex cases.

	Attending Gender	
	Female	Male
Trainee Gender (1 = Female; 0 = Male)	B= 0.0345 (0.0182)	-0.0229* (0.00955)
Fixed Effects for Attending	Yes	Yes
Fixed Effects for Training Level	Yes	Yes
Fixed Effects for Procedure	Yes	Yes
Postgraduate Year (PGY) 1	Omitted	Omitted
PGY 2	0.475*** (0.0316)	0.308*** (0.0181)
PGY 3	0.816*** (0.0306)	0.677*** (0.0163)
PGY 4	1.204*** (0.0322)	1.040*** (0.0169)
PGY 5	1.558*** (0.0345)	1.354*** (0.0175)
PGY 6	1.656*** (0.0796)	1.360*** (0.0445)
PGY 7	2.006*** (0.336)	1.605*** (0.0764)
PGY 8	1.244*** (0.294)	1.749*** (0.160)
PGY 9	1.812*** (0.319)	2.359*** (0.168)
Observations	7,343	23,088
R^2	0.391	0.353

Table S9. Autonomy gender gap by attending gender for most complex cases. Multivariable regressions predicting surgical autonomy as rated by attending was used to separately analyze female and male attending surgeons in cases rated "hardest 1/3" in complexity compared with similar cases (a subgroup analysis of the Most Complex model in Table S7). The table shows for each variable the unstandardized regression coefficient B and, in parentheses, the standard errors of B. Significance is denoted with *, **, and *** at p<0.05, p<0.01 and p<0.001 levels, respectively. The gender gap was statistically significant for female attendings in the most complex cases.

	Attending Gender	
	Female	Male
Trainee Gender (1 = Female; 0 = Male)	B= -0.142*** (0.0393)	-0.0327 (0.0184)
Fixed Effects for Attending	Yes	Yes
Fixed Effects for Training Level	Yes	Yes
Fixed Effects for Procedure	Yes	Yes
Postgraduate Year (PGY) 1	Omitted	Omitted
PGY 2	0.460*** (0.0843)	0.322*** (0.0410)
PGY 3	0.777*** (0.0777)	0.573*** (0.0370)
PGY 4	1.045*** (0.0807)	0.935*** (0.0363)
PGY 5	1.521*** (0.0820)	1.285*** (0.0366)
PGY 6	1.684*** (0.191)	1.163*** (0.0964)
PGY 7	2.042*** (0.505)	1.416*** (0.162)
PGY 8	1.693*** (0.398)	1.399*** (0.279)
PGY 9	2.238*** (0.606)	1.990*** (0.0290)
Observations	1,911	7,279
R^2	0.303	0.289

Table S10. Performance by training level. Mean performance ratings by attending surgeons are listed according to trainee postgraduate year. Performance ratings were not elicited for cases where the level of autonomy was rated as only "show and tell." SD, standard deviation.

Postgraduate Year (PGY)	Number of Assessments	Mean Performance Rating (SD)
1	4,310	2.79 (0.70)
2	5,204	3.08 (0.70)
3	7,315	3.21 (0.64)
4	9,122	3.51 (0.75)
5	7,937	3.78 (0.64)
6	853	3.73 (0.66)
7	209	3.71 (0.66)
8	684	3.71 (0.69)
9+	682	4.13 (0.67)

Chen JX, Chang EH, Deng F, et al. Autonomy in the operating room: a multicenter study of gender disparities during surgical training. *J Grad Med Educ*. 2021;13(5):666–672.
http://dx.doi.org/10.4300/JGME-D-21-00217.1

Table S11. Performance gender gap models. Multivariable regressions model surgical performance as rated by attendings based on trainee gender (models 1-4) while accounting for additional fixed effects based on attending (models 2-4), training level (models 3-4), and procedure (model 4). The table shows for each variable the unstandardized regression coefficient B and, in parentheses, the standard errors of B. Significance is denoted with *, **, and *** at p<0.05, p<0.01 and p<0.001 levels, respectively. There was no significant performance gender gap in the full model (Model 4).

	Model 1	Model 2	Model 3	Model 4
Trainee Gender (1 = Female; 0 = Male)	B= - (0.00831)	- (0.00785)	-0.00831 (0.00693)	-0.0124 (0.00673)
Fixed Effects for Attending	No	Yes	Yes	Yes
Fixed Effects for Training Level	No	No	Yes	Yes
Fixed Effects for Procedure	No	No	No	Yes
Postgraduate Year (PGY) 1			Omitted	Omitted
PGY 2			0.230*** (0.0139)	0.272*** (0.0136)
PGY 3			0.485*** (0.0127)	0.571*** (0.0124)
PGY 4			0.826*** (0.128)	0.962*** (0.0128)
PGY 5			1.124*** (0.0132)	1.293*** (0.0132)
PGY 6			1.069*** (0.0296)	1.238*** (0.0300)
PGY 7			1.180*** (0.0516)	1.399*** (0.0564)
PGY 8			1.321*** (0.0919)	1.440*** (0.0942)
PGY 9			1.834*** (0.0978)	1.967*** (0.100)
PGY 10			0.258*** (0.452)	2.398*** (0.597)
Observations	36,316	36,316	36,316	36,316
R^2	0.0013	0.0013	0.192	0.296

Table S12. Trainee self-ratings for autonomy. Multivariable regressions were used to predict surgical autonomy self-ratings based on trainee gender, accounting for fixed effects of attending rating of autonomy, attending, training level, procedure, and case complexity. The table shows for each variable the unstandardized regression coefficient B and, in parentheses, the standard errors of B. Significance is denoted with *, **, and *** at p<0.05, p<0.01 and p<0.001 levels, respectively. There was a statistically significant gender difference in trainee autonomy self-ratings.

	Trainee rating
Trainee Gender (1 = Female; 0 = Male)	B= -0.0669*** (0.0163)
Fixed Effects for Attending Rating of Autonomy	Yes
Fixed Effects for Attending	Yes
Fixed Effects for Training Level	Yes
Fixed Effects for Procedure	Yes
Fixed Effects for Case Complexity	Yes
Attending rating Zwisch level 1	Omitted
Attending rating Zwisch level 2	0.352*** (0.0200)
Attending rating Zwisch level 3	0.725*** (0.0238)
Attending rating Zwisch level 4	1.242*** (0.0288)
Observations	34,661
R^2	0.553

Table S13. Trainee self-ratings for performance. Multivariable regressions were used to predict surgical performance self-ratings based on trainee gender, accounting for fixed effects of attending rating of performance, attending, training level, procedure, and case complexity. The table shows for each variable the unstandardized regression coefficient B and, in parentheses, the standard errors of B. Significance is denoted with *, **, and *** at p<0.05, p<0.01 and p<0.001 levels, respectively. There was a statistically significant gender difference in trainee performance self-ratings.

	Trainee Rating
Trainee Gender (1 = Female; 0 = Male)	B= -0.0704*** (0.0175)
Fixed Effects for Attending Rating of Performance	Yes
Fixed Effects for Attending	Yes
Fixed Effects for Training Level	Yes
Fixed Effects for Procedure	Yes
Fixed Effects for Case Complexity	Yes
Attending rating performance level 1	Omitted
Attending rating performance level 2	-0.0435 (0.150)
Attending rating performance level 3	0.218 (0.148)
Attending rating performance level 4	0.538*** (0.148)
Attending rating performance level 5	0.756*** (0.151)
Observations	29,909
R^2	0.488

Supplementary References

1. Angrist JD, Pischke J-S. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press; 2008.

2. Abadie A, Athey S, Imbens GW, Wooldridge J. *When Should You Adjust Standard Errors for Clustering?* National Bureau of Economic Research; 2017. doi:10.3386/w24003

3. Ai C, Norton EC. Interaction terms in logit and probit models. *Economics Letters*. 2003;80(1):123-129. doi:10.1016/S0165-1765(03)00032-6

4. Chen JX, Kozin ED, Bohnen JD, et al. Assessments of otolaryngology resident operative experiences using mobile technology: a pilot study. *Otolaryngol Head Neck Surg*. 2019;161(5):939-945. doi:10.1177/0194599819868165