Supplemental Digital Appendix 1

List of Courses Analyzed by Sex and Race of Images, Academic Year 2010-2011, From a Study of Equitable Imagery in the Pre-Clinical Medical Curriculum, University of Washington School of Medicine, 2009–2011[a]

1. Histology
2. Biochemistry (Fall)
3. Biochemistry (Winter)
4. Anatomy and Embryology
5. Systems of Human Behavior
6. Introduction to Clinical Medicine 1 (Fall)
7. Introduction to Clinical Medicine 1 (Winter)
8. Introduction to Clinical Medicine 1 (Spring)
9. Medical Information and Decision Making
10. Cell Physiology
11. Musculoskeletal System
12. Immunology
13. Nervous System
14. Microbiology
15. Cardiovascular System
16. Pathology (Fall)
17. Pathology (Winter)
18. Pathology (Spring)
19. Respiratory System
20. Ethics
21. Urinary System
22. Pharmacology (Fall)
23. Pharmacology (Spring)
24. Epidemiology
25. Medicine, Health, and Society
26. Hormones/Nutrients
27. Hematology
28. Genetics
29. Problem Based Learning
30. Gastro-Intestinal Systems
31. Reproduction
32. Brain and Behavior
33. Skin System

[a] Introduction to Clinical Medicine 2 is offered in the fall, winter, and spring of 2nd year, but slides could not be analyzed as they were unavailable.

Supplemental Digital Appendix 2

Interrater reliability chart –*P* values, From a Study of Equitable Imagery in the Pre-Clinical Medical Curriculum, University of Washington School of Medicine, 2009–2011

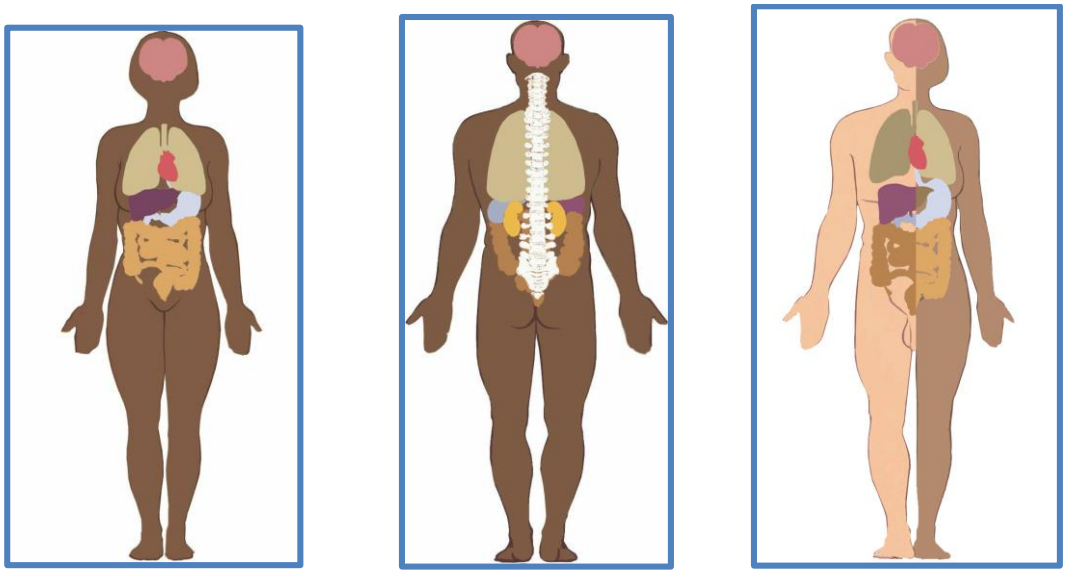
Category	C1:C2	C2:C3	C1:C3
Human	0.88	0.019[a]	0.84
Female	0.65	0.42	0.38
Male	0.80	0.25	0.85
White	0.80	0.56	0.18
POC	1.0	0.68	0.54

Abbreviations: POC indicates person of color; C, coder.

[a] Denotes significant difference. It is pertinent to mention the Bonferroni correction here, which is an adjustment made to *P* values when several dependent or independent statistical tests are being performed simultaneously on a single data set. (Source: Shaffer JP. Multiple Hypothesis Testing. *Annual Review of Psychology.* 1995;46:561–584.) Although the Bonferroni correction is generally used to reduce the chances of obtaining false-positive results (type I errors) when multiple pair wise tests are performed on a single set of data, here the correction is used to demonstrate the odds of finding a single statistically significant test among the t-tests for interrater reliability. Put another way, the probability of identifying at least one significant result simply due to chance increases as more 'hypotheses' are tested. Thus, the calculation can be performed as follows: *P* (no significant differences in all these tests) = $(1 - \alpha)^k$ and *P* (at least one significant result) = $1 – (1 - \alpha)^k = 1- (1-0.05)^{15} = 0.54$, meaning there is a 54% chance of yielding a significant result. This is not to dismiss the finding of a difference in interrater reliability in human images between rater C2 and C3, which may be real and due to differences in interpretations of images that would be considered human (for example cartoons, stick figures, etc.). The calculation is noted simply to state the drawbacks of this type of statistical testing.

Supplemental Digital Appendix 3

Examples of Didactic Images for Pre-Clinical Education that Could be Used in the Future by a School of Medicine for More Equal Representation of Sex and Race, From a Study of Equitable Imagery in the Pre-Clinical Medical Curriculum, University of Washington School of Medicine, 2009–2011



Images created by Rick Fisher, rickfisherart.com
Exclusively created for the authors, no previous publication