A Evaluation Information

A.1 Metrics

Precision. To evaluate the change that image corruptions have to face detection systems, we measure the precision of the corrupted images while using the detections from the clean image as ground truth. While this approach obviates the need for real ground truth bounding boxes, it is also a principled measurement strategy for our main research question. Since we are primarily interested in how the system is affected by the corruption, this metric is superior to using real ground truth bounding boxes. This follows because we're interested in isolating the change in a system under a corruption which is exactly what this method measures.

To compute precision, we first observe the face detections on each clean image. After subsequently observing the face detection of a corrupted version of the clean image, we compute the image-level precision and recall for the corrupted image while using whatever the clean image's detections were as ground truth.

A.2 Image Counts

For each dataset, we selected no more than 1500 images from any intersectional group. The final tallies of how many images from each group can be found in Tables 2, 3, 4, and 5.

A.3 Corruption information

We evaluate 15 corruptions from Hendrycks and Dietterich [35]: Gaussian noise, shot noise, impulse noise, defocus blur, glass blur, motion blur, zoom blur, snow, frost, fog, brightness, contrast, elastic transforms, pixelation, and jpeg compressions. Each corruption is described in the Hendrycks and Dietterich [35] paper as follows:

The first corruption type is Gaussian noise. This corruption can appear in low-lighting conditions. Shot noise, also called Poisson noise, is electronic noise caused by the discrete nature of light itself. Impulse noise is a color analogue of salt-and-pepper noise and can be caused by bit errors. Defocus blur occurs when an image is out of focus. Frosted Glass Blur appears with "frosted glass" windows or panels. Motion blur appears when a camera is moving quickly. Zoom blur occurs when a camera moves toward an object rapidly. Snow is a visually obstructive form of precipitation. Frost forms when lenses or windows are coated with ice crystals. Fog shrouds objects and is rendered with the diamond-square algorithm. Brightness varies with daylight intensity. Contrast can be high or low depending on lighting conditions and the photographed object's color. Elastic transformations stretch or contract small image regions. Pixelation occurs when upsampling a lowresolution image. JPEG is a lossy image compression format which introduces compression artifacts.

The specific parameters for each corruption can be found in the project's github at the corruptions file: `https://github.com/dooleys/Robustness-Disparities-in-Commercial-Face-Detection/blob/main/code/imagenet_c_big/corruptions.py`.

B API Parameters

For the AWS DetectFaces API,[2] we selected to have all facial attributes returned. The Azure Face API[3] allows the user to select one of three detection models. We chose model `detection_03` as it was their most recently released model (February 2021) and was described to have the highest performance on small, side, and blurry faces, since it aligns with our benchmark intention. This model does not return age or gender estimates (though model `detection_01` does).

These experiments were conducted in July 2021.

C Benchmarks Costs

A total breakdown of costs for this benchmark can be found in Table 6.

[2] `https://docs.aws.amazon.com/rekognition/latest/dg/API_DetectFaces.html`

[3] `https://westus.dev.cognitive.microsoft.com/docs/services/563879b61984550e40cbbe8d/operations/563879b61984550f30395236`

Table 1: The best and worst performing perturbations for each dataset and model.

		AWS	Azure	GCP	MogFace	TinaFace	Yolo5Face
Adience	Best	Brightness	Brightness	Pixelate	Brightness	Brightness	Brightness
	Worst	Impulse Noise	Shot Noise	Snow	Impulse Noise	Impulse Noise	Impulse Noise
CCD	Best	Glass Blur	Glass Blur	Glass Blur	Glass Blur	Glass Blur	Glass Blur
	Worst	Zoom Blur	Zoom Blur	Zoom Blur	Zoom Blur	Zoom Blur	Zoom Blur
MIAP	Best	Brightness	Glass Blur	JPEG Compression	Brightness	Brightness	Brightness
	Worst	Zoom Blur	Zoom Blur	Zoom Blur	Zoom Blur	Zoom Blur	Zoom Blur
UTKFace	Best	Brightness	Brightness	JPEG Compression	Brightness	Brightness	Brightness
	Worst	Elastic Transform	Elastic Transform	Elastic Transform	Elastic Transform	Impulse Noise	Shot Noise

Table 2: Adience Dataset Counts

Age	Gender	Count
0-2	Female	684
	Male	716
3-7	Female	1232
	Male	925
8-14	Female	1353
	Male	933
15-24	Female	1047
	Male	742
25-35	Female	1500
	Male	1500
36-45	Female	1078
	Male	1412
46-59	Female	436
	Male	466
60+	Female	428
	Male	467

D Statistical Significance Regressions for Average Precision

D.1 Main Tables

AP p-values for pairwise Wilcox test with Bonferroni correction for model on the Adience dataset can befound in Table 7

AP p-values for pairwise Wilcox test with Bonferroni correction for model on the CCD dataset can befound in Table 8

AP p-values for pairwise Wilcox test with Bonferroni correction for model on the MIAP dataset can befound in Table 9

AP p-values for pairwise Wilcox test with Bonferroni correction for model on the UTK dataset can befound in Table 10

D.2 AP — Corruption Comparison Claims

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on AWS and Adience can be found in Table 11

Table 3: CCD Dataset Counts

Lighting	Gender	Skin	Age	Count
Bright	Female	Dark	19-45	1500
			45-64	1500
			65+	547
		Light	19-45	1500
			45-64	1500
			65+	653
	Male	Dark	19-45	1500
			45-64	1500
			65+	384
		Light	19-45	1500
			45-64	1500
			65+	695
	Other	Dark	19-45	368
			45-64	168
			65+	12
		Light	19-45	244
			45-64	49
Dim	Female	Dark	19-45	1500
			45-64	670
			65+	100
		Light	19-45	642
			45-64	314
			65+	131
	Male	Dark	19-45	1500
			45-64	387
			65+	48
		Light	19-45	485
			45-64	299
			65+	123
	Other	Dark	19-45	57
			45-64	26
			65+	3
		Light	19-45	27
			45-64	12

Table 4: MIAP Dataset Counts

AgePresentation	GenderPresentation	Count
Young	Unknown	1500
Middle	Predominantly Feminine	1500
	Predominantly Masculine	1500
	Unknown	561
Older	Predominantly Feminine	209
	Predominantly Masculine	748
	Unknown	24
Unknown	Predominantly Feminine	250
	Predominantly Masculine	402
	Unknown	1500

Table 5: UTKFace Dataset Counts

Age	Gender	Race	Count
0-18	Female	Asian	555
		Black	161
		Indian	350
		Others	338
		White	987
	Male	Asian	586
		Black	129
		Indian	277
		Others	189
		White	955
19-45	Female	Asian	1273
		Black	1500
		Indian	1203
		Others	575
		White	1500
	Male	Asian	730
		Black	1499
		Indian	1264
		Others	477
		White	1500
45-64	Female	Asian	39
		Black	206
		Indian	146
		Others	22
		White	802
	Male	Asian	180
		Black	401
		Indian	653
		Others	97
		White	1500
65+	Female	Asian	75
		Black	78
		Indian	43
		Others	10
		White	712
	Male	Asian	148
		Black	166
		Indian	91
		Others	5
		White	682

Table 6: Total Costs of Benchmark

Category	Cost
Azure Face Service	$4,270.58
AWS Rekognition	$4,270.66
Google Cloud Platform	$7,230.47
S3	$1,003.83
EC2	$475.77
Tax	$256.24
Total	$17,507.55

Table 7: AP. Pairwise Wilcox test with Bonferroni correction for model on Adience

	AWS	Azure	GCP	MogFace	TinaFace
Azure	0				
GCP	0	0			
MogFace	0	0	0		
TinaFace	0	0	0	0	
Yolov5	0	0	0	0	0

Table 8: AP. Pairwise Wilcox test with Bonferroni correction for model on CCD

	AWS	Azure	GCP	MogFace	TinaFace
Azure	0				
GCP	0	0			
MogFace	0	0.071	0		
TinaFace	0	0	0	0	
Yolov5	0	0	0	0	0

Table 9: AP. Pairwise Wilcox test with Bonferroni correction for model on MIAP

	AWS	Azure	GCP	MogFace	TinaFace
Azure	0				
GCP	0	0			
MogFace	0	0	0		
TinaFace	0	0	0	0	
Yolov5	0	0	0	0	0

Table 10: AP. Pairwise Wilcox test with Bonferroni correction for model on UTK

	AWS	Azure	GCP	MogFace	TinaFace
Azure	0				
GCP	0	0			
MogFace	0	0	0		
TinaFace	0	0	0	0	
Yolov5	0	0	0	0	0

Table 11: AP. Pairwise Wilcox test with Bonferroni correction for corruption on AWS and Adience

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0.099													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0.779									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0.00000	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on Azure and Adience can be found in Table 12

Table 12: AP. Pairwise Wilcox test with Bonferroni correction for corruption on Azure and Adience

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0.958	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0.0003						
fog	0	0	0	0.008	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on GCP and Adience can be found in Table 13

Table 13: AP. Pairwise Wilcox test with Bonferroni correction for corruption on GCP and Adience

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0	0.0005												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0.278	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on MogFace and Adience can be found in Table 14

Table 14: AP. Pairwise Wilcox test with Bonferroni correction for corruption on MogFace and Adience

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0.0001	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0.120	0			
elastic-transform	0	0	0	0	0	0	0.034	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on TinaFace and Adience can be found in Table 15

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on Yolov5 and Adience can be found in Table 16

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on AWS and CCD can be found in Table 17

Table 15: AP. Pairwise Wilcox test with Bonferroni correction for corruption on TinaFace and Adience

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0.00001													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0.0001	0	0	0	
jpeg-compression	0	0	0	0	0	0.047	0	0	0	0	0	0	0	0

Table 16: AP. Pairwise Wilcox test with Bonferroni correction for corruption on Yolov5 and Adience

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0.004													
impulse-noise	0	0.00000												
defocus-blur	0	0	0											
glass-blur	0	0	0	0.005										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0.00000	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0.643		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 17: AP. Pairwise Wilcox test with Bonferroni correction for corruption on AWS and CCD

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00001

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on Azure and CCD can be found in Table 18

Table 18: AP. Pairwise Wilcox test with Bonferroni correction for corruption on Azure and CCD

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0.00000	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on GCP and CCD can be found in Table 19

Table 19: AP. Pairwise Wilcox test with Bonferroni correction for corruption on GCP and CCD

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0.00000	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on MogFace and CCD can be found in Table 20

Table 20: AP. Pairwise Wilcox test with Bonferroni correction for corruption on MogFace and CCD

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0.012	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on TinaFace and CCD can be found in Table 21

Table 21: AP. Pairwise Wilcox test with Bonferroni correction for corruption on TinaFace and CCD

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0.00000	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0.016	0	0.065	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00000

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on Yolov5 and CCD can be found in Table 22

Table 22: AP. Pairwise Wilcox test with Bonferroni correction for corruption on Yolov5 and CCD

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0.822

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on AWS and MIAP can be found in Table 23

Table 23: AP. Pairwise Wilcox test with Bonferroni correction for corruption on AWS and MIAP

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0.018													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0.00000	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0.009	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on Azure and MIAP can be found in Table 24

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on GCP and MIAP can be found in Table 25

Table 24: AP. Pairwise Wilcox test with Bonferroni correction for corruption on Azure and MIAP

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0.211													
impulse-noise	0.913	0.170												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0.001	0	0.00000						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0.203	0.730	0.061	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0.068

Table 25: AP. Pairwise Wilcox test with Bonferroni correction for corruption on GCP and MIAP

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0.123													
impulse-noise	0.131	0.963												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0.018	0.450	0.309	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0.006						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0.001	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0.492	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on MogFace and MIAP can be found in Table 26

Table 26: AP. Pairwise Wilcox test with Bonferroni correction for corruption on MogFace and MIAP

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0.247													
impulse-noise	0.024	0.001												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0.575	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on TinaFace and MIAP can be found in Table 27

Table 27: AP. Pairwise Wilcox test with Bonferroni correction for corruption on TinaFace and MIAP

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0.571													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0.215	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0.0004	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on Yolov5 and MIAP can be found in Table 28

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on AWS and UTK can be found in Table 29

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on Azure and UTK can be found in Table 30

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on GCP and UTK can be found in Table 31

Table 28: AP. Pairwise Wilcox test with Bonferroni correction for corruption on Yolov5 and MIAP

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0.002													
impulse-noise	0.00000	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0.013	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0.00003	0	0			
elastic-transform	0.014	0	0.007	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0.014	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 29: AP. Pairwise Wilcox test with Bonferroni correction for corruption on AWS and UTK

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0.181	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0.756

Table 30: AP. Pairwise Wilcox test with Bonferroni correction for corruption on Azure and UTK

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0.272	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0.272						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0.084	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 31: AP. Pairwise Wilcox test with Bonferroni correction for corruption on GCP and UTK

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0.003										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0.357	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0.001	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on MogFace and UTK can be found in Table 32

Table 32: AP. Pairwise Wilcox test with Bonferroni correction for corruption on MogFace and UTK

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on TinaFace and UTK can be found in Table 33

Table 33: AP. Pairwise Wilcox test with Bonferroni correction for corruption on TinaFace and UTK

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0.031	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for corruption on Yolov5 and UTK can be found in Table 34

Table 34: AP. Pairwise Wilcox test with Bonferroni correction for corruption on Yolov5 and UTK

	gaussian-noise	shot-noise	impulse-noise	defocus-blur	glass-blur	motion-blur	zoom-blur	snow	frost	fog	brightness	contrast	elastic-transform	pixelate
shot-noise	0													
impulse-noise	0	0												
defocus-blur	0	0	0											
glass-blur	0	0	0	0										
motion-blur	0	0	0	0	0									
zoom-blur	0	0	0	0	0	0								
snow	0	0	0	0	0	0	0							
frost	0	0	0	0	0	0	0	0						
fog	0	0	0	0	0	0	0	0	0					
brightness	0	0	0	0	0	0	0	0	0	0				
contrast	0	0	0	0	0	0	0	0	0	0	0			
elastic-transform	0	0	0	0	0	0	0	0	0	0	0	0		
pixelate	0	0	0	0	0	0	0	0	0	0	0	0	0	
jpeg-compression	0	0	0	0	0	0	0	0	0	0	0	0.010	0	0

D.3 AP — Age Comparison Claims

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on AWS and Adience can be found in Table 35

Table 35: AP. Pairwise Wilcox test with Bonferroni correction for Age on AWS and Adience

	0-2	3-7	8-14	15-24	25-35	36-45	46-59
3-7	0						
8-14	0	0					
15-24	0	0	0				
25-35	0	0	0	0			
36-45	0	0	0	0	0		
46-59	0	0.00000	0	0	0	0	
60+	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on Azure and Adience can be found in Table 36

Table 36: AP. Pairwise Wilcox test with Bonferroni correction for Age on Azure and Adience

	0-2	3-7	8-14	15-24	25-35	36-45	46-59
3-7	0						
8-14	0	0					
15-24	0	0	0				
25-35	0	0	0	0			
36-45	0	0.00000	0	0	0.00000		
46-59	0	0.118	0	0	0	0	
60+	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on GCP and Adience can be found in Table 37

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on MogFace and Adience can be found in Table 38

Table 37: AP. Pairwise Wilcox test with Bonferroni correction for Age on GCP and Adience

	0-2	3-7	8-14	15-24	25-35	36-45	46-59
3-7	0						
8-14	0	0					
15-24	0	0.00004	0				
25-35	0	0	0	0.134			
36-45	0.008	0	0	0	0		
46-59	0	0	0	0	0	0	
60+	0	0	0	0	0	0	0.003

Table 38: AP. Pairwise Wilcox test with Bonferroni correction for Age on MogFace and Adience

	0-2	3-7	8-14	15-24	25-35	36-45	46-59
3-7	0						
8-14	0	0					
15-24	0	0	0.945				
25-35	0	0	0.001	0.003			
36-45	0	0	0	0	0		
46-59	0	0.524	0	0	0	0	
60+	0.198	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on TinaFace and Adience can be found in Table 39

Table 39: AP. Pairwise Wilcox test with Bonferroni correction for Age on TinaFace and Adience

	0-2	3-7	8-14	15-24	25-35	36-45	46-59
3-7	0						
8-14	0	0					
15-24	0	0	0				
25-35	0	0	0	0			
36-45	0	0	0.0001	0	0		
46-59	0	0	0.005	0	0	0	
60+	0.010	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on Yolov5 and Adience can be found in Table 40

Table 40: AP. Pairwise Wilcox test with Bonferroni correction for Age on Yolov5 and Adience

	0-2	3-7	8-14	15-24	25-35	36-45	46-59
3-7	0.226						
8-14	0	0					
15-24	0	0.00000	0.00000				
25-35	0.00000	0.0001	0	0.049			
36-45	0	0	0	0	0		
46-59	0	0	0.0002	0	0	0.0001	
60+	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on AWS and CCD can be found in Table 41

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on Azure and CCD can be found in Table 42

Table 41: AP. Pairwise Wilcox test with Bonferroni correction for Age on AWS and CCD

	19-45	45-64
45-64	0	
65+	0	0

Table 42: AP. Pairwise Wilcox test with Bonferroni correction for Age on Azure and CCD

	19-45	45-64
45-64	0	
65+	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on GCP and CCD can be found in Table 43

Table 43: AP. Pairwise Wilcox test with Bonferroni correction for Age on GCP and CCD

	19-45	45-64
45-64	0	
65+	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on MogFace and CCD can be found in Table 44

Table 44: AP. Pairwise Wilcox test with Bonferroni correction for Age on MogFace and CCD

	19-45	45-64
45-64	0	
65+	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on TinaFace and CCD can be found in Table 45

Table 45: AP. Pairwise Wilcox test with Bonferroni correction for Age on TinaFace and CCD

	19-45	45-64
45-64	0	
65+	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on Yolov5 and CCD can be found in Table 46

Table 46: AP. Pairwise Wilcox test with Bonferroni correction for Age on Yolov5 and CCD

	19-45	45-64
45-64	0	
65+	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on AWS and MIAP can be found in Table 47

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on Azure and MIAP can be found in Table 48

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on GCP and MIAP can be found in Table 49

Table 47: AP. Pairwise Wilcox test with Bonferroni correction for Age on AWS and MIAP

	Young	Middle	Older
Middle	0		
Older	0	0	
Unknown	0	0	0

Table 48: AP. Pairwise Wilcox test with Bonferroni correction for Age on Azure and MIAP

	Young	Middle	Older
Middle	0		
Older	0	0	
Unknown	0	0.00000	0

Table 49: AP. Pairwise Wilcox test with Bonferroni correction for Age on GCP and MIAP

	Young	Middle	Older
Middle	0		
Older	0	0	
Unknown	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on MogFace and MIAP can be found in Table 50

Table 50: AP. Pairwise Wilcox test with Bonferroni correction for Age on MogFace and MIAP

	Young	Middle	Older
Middle	0		
Older	0	0	
Unknown	0	0.00000	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on TinaFace and MIAP can be found in Table 51

Table 51: AP. Pairwise Wilcox test with Bonferroni correction for Age on TinaFace and MIAP

	Young	Middle	Older
Middle	0		
Older	0	0	
Unknown	0	0.001	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on Yolov5 and MIAP can be found in Table 52

Table 52: AP. Pairwise Wilcox test with Bonferroni correction for Age on Yolov5 and MIAP

	Young	Middle	Older
Middle	0		
Older	0	0	
Unknown	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on AWS and UTK can be found in Table 53

Table 53: AP. Pairwise Wilcox test with Bonferroni correction for Age on AWS and UTK

	0-18	19-45	45-64
19-45	0		
45-64	0	0	
65+	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on Azure and UTK can be found in Table 54

Table 54: AP. Pairwise Wilcox test with Bonferroni correction for Age on Azure and UTK

	0-18	19-45	45-64
19-45	0		
45-64	0	0.570	
65+	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on GCP and UTK can be found in Table 55

Table 55: AP. Pairwise Wilcox test with Bonferroni correction for Age on GCP and UTK

	0-18	19-45	45-64
19-45	0		
45-64	0	0	
65+	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on MogFace and UTK can be found in Table 56

Table 56: AP. Pairwise Wilcox test with Bonferroni correction for Age on MogFace and UTK

	0-18	19-45	45-64
19-45	0		
45-64	0	0	
65+	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on TinaFace and UTK can be found in Table 57

Table 57: AP. Pairwise Wilcox test with Bonferroni correction for Age on TinaFace and UTK

	0-18	19-45	45-64
19-45	0		
45-64	0.00000	0	
65+	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Age on Yolov5 and UTK can be found in Table 58

D.4 AP — Gender Comparison Claims

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on AWS and Adience can be found in Table 59

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on Azure and Adience can be found in Table 60

Table 58: AP. Pairwise Wilcox test with Bonferroni correction for Age on Yolov5 and UTK

	0-18	19-45	45-64
19-45	0		
45-64	0	0	
65+	0	0	0

Table 59: AP. Pairwise Wilcox test with Bonferroni correction for Gender on AWS and Adience

	Female
Male	0

Table 60: AP. Pairwise Wilcox test with Bonferroni correction for Gender on Azure and Adience

	Female
Male	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on GCP and Adience can be found in Table 61

Table 61: AP. Pairwise Wilcox test with Bonferroni correction for Gender on GCP and Adience

	Female
Male	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on MogFace and Adience can be found in Table 62

Table 62: AP. Pairwise Wilcox test with Bonferroni correction for Gender on MogFace and Adience

	Female
Male	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on TinaFace and Adience can be found in Table 63

Table 63: AP. Pairwise Wilcox test with Bonferroni correction for Gender on TinaFace and Adience

	Female
Male	0.203

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on Yolov5 and Adience can be found in Table 64

Table 64: AP. Pairwise Wilcox test with Bonferroni correction for Gender on Yolov5 and Adience

	Female
Male	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on AWS and CCD can be found in Table 65

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on Azure and CCD can be found in Table 66

Table 65: AP. Pairwise Wilcox test with Bonferroni correction for Gender on AWS and CCD

	Female	Male
Male	0	
Other	0.680	0

Table 66: AP. Pairwise Wilcox test with Bonferroni correction for Gender on Azure and CCD

	Female	Male
Male	0	
Other	0.171	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on GCP and CCD can be found in Table 67

Table 67: AP. Pairwise Wilcox test with Bonferroni correction for Gender on GCP and CCD

	Female	Male
Male	0	
Other	0.003	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on MogFace and CCD can be found in Table 68

Table 68: AP. Pairwise Wilcox test with Bonferroni correction for Gender on MogFace and CCD

	Female	Male
Male	0	
Other	0.806	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on TinaFace and CCD can be found in Table 69

Table 69: AP. Pairwise Wilcox test with Bonferroni correction for Gender on TinaFace and CCD

	Female	Male
Male	0	
Other	0.740	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on Yolov5 and CCD can be found in Table 70

Table 70: AP. Pairwise Wilcox test with Bonferroni correction for Gender on Yolov5 and CCD

	Female	Male
Male	0	
Other	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on AWS and MIAP can be found in Table 71

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on Azure and MIAP can be found in Table 72

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on GCP and MIAP can be found in Table 73

Table 71: AP. Pairwise Wilcox test with Bonferroni correction for Gender on AWS and MIAP

	Predominantly Feminine	Predominantly Masculine
Predominantly Masculine	0	
Unknown	0	0

Table 72: AP. Pairwise Wilcox test with Bonferroni correction for Gender on Azure and MIAP

	Predominantly Feminine	Predominantly Masculine
Predominantly Masculine	0	
Unknown	0	0

Table 73: AP. Pairwise Wilcox test with Bonferroni correction for Gender on GCP and MIAP

	Predominantly Feminine	Predominantly Masculine
Predominantly Masculine	0	
Unknown	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on MogFace and MIAP can be found in Table 74

Table 74: AP. Pairwise Wilcox test with Bonferroni correction for Gender on MogFace and MIAP

	Predominantly Feminine	Predominantly Masculine
Predominantly Masculine	0	
Unknown	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on TinaFace and MIAP can be found in Table 75

Table 75: AP. Pairwise Wilcox test with Bonferroni correction for Gender on TinaFace and MIAP

	Predominantly Feminine	Predominantly Masculine
Predominantly Masculine	0	
Unknown	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on Yolov5 and MIAP can be found in Table 76

Table 76: AP. Pairwise Wilcox test with Bonferroni correction for Gender on Yolov5 and MIAP

	Predominantly Feminine	Predominantly Masculine
Predominantly Masculine	0	
Unknown	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on AWS and UTK can be found in Table 77

Table 77: AP. Pairwise Wilcox test with Bonferroni correction for Gender on AWS and UTK

	Female
Male	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on Azure and UTK can be found in Table 78

Table 78: AP. Pairwise Wilcox test with Bonferroni correction for Gender on Azure and UTK

	Female
Male	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on GCP and UTK can be found in Table 79

Table 79: AP. Pairwise Wilcox test with Bonferroni correction for Gender on GCP and UTK

	Female
Male	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on MogFace and UTK can be found in Table 80

Table 80: AP. Pairwise Wilcox test with Bonferroni correction for Gender on MogFace and UTK

	Female
Male	0.0001

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on TinaFace and UTK can be found in Table 81

Table 81: AP. Pairwise Wilcox test with Bonferroni correction for Gender on TinaFace and UTK

	Female
Male	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Gender on Yolov5 and UTK can be found in Table 82

Table 82: AP. Pairwise Wilcox test with Bonferroni correction for Gender on Yolov5 and UTK

	Female
Male	0

D.5 AP — Skin Type Comparison Claims

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type on AWS and CCD can be found in Table 83

Table 83: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type on AWS and CCD

	Light Fitz
Dark Fitz	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type on Azure and CCD can be found in Table 84

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type on GCP and CCD can be found in Table 85

Table 84: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type on Azure and CCD

	Light Fitz
Dark Fitz	0

Table 85: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type on GCP and CCD

	Light Fitz
Dark Fitz	0

Table 86: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type on MogFace and CCD

	Light Fitz
Dark Fitz	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type on MogFace and CCD can be found in Table 86

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type on TinaFace and CCD can be found in Table 87

Table 87: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type on TinaFace and CCD

	Light Fitz
Dark Fitz	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type on Yolov5 and CCD can be found in Table 88

Table 88: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type on Yolov5 and CCD

	Light Fitz
Dark Fitz	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Lighting on AWS and CCD can be found in Table 89

Table 89: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Lighting on AWS and CCD

	Dark Fitz+Bright	Dark Fitz+Dim	Light Fitz+Bright
Dark Fitz+Dim	0		
Light Fitz+Bright	0	0	
Light Fitz+Dim	0	0.567	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Lighting on Azure and CCD can be found in Table 90

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Lighting on GCP and CCD can be found in Table 91

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Lighting on MogFace and CCD can be found in Table 92

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Lighting on TinaFace and CCD can be found in Table 93

Table 90: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Lighting on Azure and CCD

	Dark Fitz+Bright	Dark Fitz+Dim	Light Fitz+Bright
Dark Fitz+Dim	0		
Light Fitz+Bright	0	0	
Light Fitz+Dim	0	0.076	0

Table 91: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Lighting on GCP and CCD

	Dark Fitz+Bright	Dark Fitz+Dim	Light Fitz+Bright
Dark Fitz+Dim	0		
Light Fitz+Bright	0	0	
Light Fitz+Dim	0	0	0

Table 92: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Lighting on MogFace and CCD

	Dark Fitz+Bright	Dark Fitz+Dim	Light Fitz+Bright
Dark Fitz+Dim	0		
Light Fitz+Bright	0	0	
Light Fitz+Dim	0	0.316	0

Table 93: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Lighting on TinaFace and CCD

	Dark Fitz+Bright	Dark Fitz+Dim	Light Fitz+Bright
Dark Fitz+Dim	0		
Light Fitz+Bright	0	0	
Light Fitz+Dim	0	0.004	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Lighting on Yolov5 and CCD can be found in Table 94

Table 94: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Lighting on Yolov5 and CCD

	Dark Fitz+Bright	Dark Fitz+Dim	Light Fitz+Bright
Dark Fitz+Dim	0		
Light Fitz+Bright	0	0	
Light Fitz+Dim	0	0.00004	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Age and Gender on AWS and CCD can be found in Table 95

Table 95: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Age and Gender on AWS and CCD

	Dark Fitz + 19-45 + Female	Dark Fitz + 19-45 + Male	Dark Fitz + 19-45 + Other	Dark Fitz + 45-64 + Female	Dark Fitz + 45-64 + Male	Dark Fitz + 45-64 + Other	Dark Fitz + 65+ + Female	Dark Fitz + 65+ + Male	Dark Fitz + 65+ + Other	Light Fitz + 19-45 + Female	Light Fitz + 19-45 + Male	Light Fitz + 19-45 + Other	Light Fitz + 45-64 + Female	Light Fitz + 45-64 + Male	Light Fitz + 45-64 + Other	Light Fitz + 65+ + Female
Dark Fitz + 19-45 + Male	0															
Dark Fitz + 19-45 + Other	0.026	0.00000														
Dark Fitz + 45-64 + Female	0.002	0	0.760													
Dark Fitz + 45-64 + Male	0	0.00000	0	0												
Dark Fitz + 45-64 + Other	0.061	0.763	0.004	0.002	0.031											
Dark Fitz + 65+ + Female	0.00000	0.125	0.00000	0	0.054	0.268										
Dark Fitz + 65+ + Male	0	0.002	0	0	0.004	0.029	0.125									
Dark Fitz + 65+ + Other	0	0	0	0	0	0	0	0								
Light Fitz + 19-45 + Female	0	0	0	0	0	0	0	0	0							
Light Fitz + 19-45 + Male	0	0	0	0	0	0	0	0	0	0.021						
Light Fitz + 19-45 + Other	0	0	0.00005	0.00000	0	0.041	0	0	0	0.191	0.006					
Light Fitz + 45-64 + Female	0.016	0	0.003	0.001	0.008	0.168	0	0	0	0	0	0.00000				
Light Fitz + 45-64 + Male	0.000	0.016	0.000	0.000	0	0.530	0.002	0.00005	0	0	0	0	0.00000			
Light Fitz + 45-64 + Other	0.006	0.993	0.067	0.042	0.00000	0.691	0.00051	0.00000	0	0.212	0.070	0.007	0.027			
Light Fitz + 65+ + Female	0.00000	0.107	0.00000	0	0.012	0.241	0.769	0.063	0	0	0	0	0	0.093	0.00005	
Light Fitz + 65+ + Male	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Age and Gender on Azure and CCD can be found in Table 96

Table 96: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Age and Gender on Azure and CCD

	Dark Fitz + 19-45 + Female	Dark Fitz + 19-45 + Male	Dark Fitz + 19-45 + Other	Dark Fitz + 45-64 + Female	Dark Fitz + 45-64 + Male	Dark Fitz + 45-64 + Other	Dark Fitz + 65+ + Female	Dark Fitz + 65+ + Male	Dark Fitz + 65+ + Other	Light Fitz + 19-45 + Female	Light Fitz + 19-45 + Male	Light Fitz + 19-45 + Other	Light Fitz + 45-64 + Female	Light Fitz + 45-64 + Male	Light Fitz + 45-64 + Other	Light Fitz + 65+ + Female
Dark Fitz + 19-45 + Male	0															
Dark Fitz + 19-45 + Other	0.947	0														
Dark Fitz + 45-64 + Female	0.130	0.289	0													
Dark Fitz + 45-64 + Male	0	0.00000	0	0												
Dark Fitz + 45-64 + Other	0.091	0.0002	0.135	0.236	0.00000											
Dark Fitz + 65+ + Female	0.00000	0.00002	0.0002	0.00003	0	0.236										
Dark Fitz + 65+ + Male	0	0.052	0	0	0.516	0.00001	0.00000									
Dark Fitz + 65+ + Other	0	0	0	0	0	0	0	0								
Light Fitz + 19-45 + Female	0	0	0	0	0	0	0	0	0							
Light Fitz + 19-45 + Male	0	0	0	0	0	0	0	0	0	0.012						
Light Fitz + 19-45 + Other	0.062	0	0.342	0.016	0	0.039	0.00000	0	0	0	0.00000					
Light Fitz + 45-64 + Female	0.902	0	0.990	0.143	0	0.078	0.00000	0	0	0	0	0.088				
Light Fitz + 45-64 + Male	0	0	0.00001	0	0	0.202	0.947	0.00000	0	0.023	0.004	0.00000	0	0		
Light Fitz + 45-64 + Other	0	0	0	0	0	0	0.27	0	0	0	0	0	0	0.067	0	
Light Fitz + 65+ + Female	0	0.141	0	0	0.00000	0.007	0.62	0.00	0	0	0	0	0	0	0	0.010
Light Fitz + 65+ + Male	0	0.092	0	0	0.104	0.00001	0.00000	0.589	0	0	0	0	0	0	0	

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Age and Gender on GCP and CCD can be found in Table 97

Table 97: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Age and Gender on GCP and CCD

	Dark Fitz + 19-45 + Female	Dark Fitz + 19-45 + Male	Dark Fitz + 19-45 + Other	Dark Fitz + 45-64 + Female	Dark Fitz + 45-64 + Male	Dark Fitz + 45-64 + Other	Dark Fitz + 65+ + Female	Dark Fitz + 65+ + Male	Dark Fitz + 65+ + Other	Light Fitz + 19-45 + Female	Light Fitz + 19-45 + Male	Light Fitz + 19-45 + Other	Light Fitz + 45-64 + Female	Light Fitz + 45-64 + Male	Light Fitz + 45-64 + Other	Light Fitz + 65+ + Female
Dark Fitz + 19-45 + Male	0															
Dark Fitz + 19-45 + Other	0.003	0														
Dark Fitz + 45-64 + Female	0.0010	0.553	0													
Dark Fitz + 45-64 + Male	0	0.00000	0	0												
Dark Fitz + 45-64 + Other	0.0004	0	0.169	0.050	0											
Dark Fitz + 65+ + Female	0	0.009	0	0	0.340	0										
Dark Fitz + 65+ + Male	0	0.526	0	0	0.0003	0	0.012									
Dark Fitz + 65+ + Other	0	0	0	0	0	0	0	0								
Light Fitz + 19-45 + Female	0	0	0	0	0	0	0	0	0							
Light Fitz + 19-45 + Male	0	0	0	0	0	0.00000	0	0	0	0						
Light Fitz + 19-45 + Other	0	0	0	0	0	0	0	0	0	0.473	0.002					
Light Fitz + 45-64 + Female	0	0	0.0002	0	0	0.291	0	0	0	0	0	0				
Light Fitz + 45-64 + Male	0.239	0	0.030	0.008	0	0.032	0	0	0	0	0	0	0			
Light Fitz + 45-64 + Other	0	0	0.00000	0.00000	0	0.0001	0	0	0	0.654	0.153	0.947	0.0002	0		
Light Fitz + 65+ + Female	0.763	0	0.001	0.0001	0	0.0001	0	0	0	0	0	0	0	0.243	0	
Light Fitz + 65+ + Male	0.269	0	0.001	0.0001	0	0.0001	0	0	0	0	0	0	0	0.058	0	0.459

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Age and Gender on MogFace and CCD can be found in Table 98

Table 98: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Age and Gender on MogFace and CCD

	Dark Fitz + 19-45 + Female	Dark Fitz + 19-45 + Male	Dark Fitz + 19-45 + Other	Dark Fitz + 45-64 + Female	Dark Fitz + 45-64 + Male	Dark Fitz + 45-64 + Other	Dark Fitz + 65+ + Female	Dark Fitz + 65+ + Male	Dark Fitz + 65+ + Other	Light Fitz + 19-45 + Female	Light Fitz + 19-45 + Male	Light Fitz + 19-45 + Other	Light Fitz + 45-64 + Female	Light Fitz + 45-64 + Male	Light Fitz + 45-64 + Other	Light Fitz + 65+ + Female
Dark Fitz + 19-45 + Male	0															
Dark Fitz + 19-45 + Other	0.010	0														
Dark Fitz + 45-64 + Female	0.475	0	0.003													
Dark Fitz + 45-64 + Male	0	0.00001	0	0												
Dark Fitz + 45-64 + Other	0.00004	0.085	0.00000	0.0002	0.0004											
Dark Fitz + 65+ + Female	0.0002	0	0.00000	0.001	0	0.100										
Dark Fitz + 65+ + Male	0	0.0003	0	0	0.221	0.0002	0									
Dark Fitz + 65+ + Other	0	0	0	0	0	0	0	0								
Light Fitz + 19-45 + Female	0	0	0	0	0	0	0	0	0							
Light Fitz + 19-45 + Male	0	0	0.0001	0	0	0	0	0	0	0						
Light Fitz + 19-45 + Other	0	0	0.00000	0	0	0	0	0	0	0.011	0.026					
Light Fitz + 45-64 + Female	0	0	0	0	0	0	0	0	0	0	0.00000	0.00001				
Light Fitz + 45-64 + Male	0	0.732	0	0	0.004	0.030	0	0.005	0	0	0	0	0			
Light Fitz + 45-64 + Other	0	0	0.00000	0	0	0	0	0	0	0.279	0.001	0.000	0.00003	0		
Light Fitz + 65+ + Female	0.0003	0	0.00000	0.003	0	0.053	0.718	0	0	0	0	0	0	0	0	
Light Fitz + 65+ + Male	0	0	0	0	0	0	0	0.00000	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Age and Gender on TinaFace and CCD can be found in Table 99

Table 99: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Age and Gender on TinaFace and CCD

	Dark Fitz + 19-45 + Female	Dark Fitz + 19-45 + Male	Dark Fitz + 19-45 + Other	Dark Fitz + 45-64 + Female	Dark Fitz + 45-64 + Male	Dark Fitz + 45-64 + Other	Dark Fitz + 65+ + Female	Dark Fitz + 65+ + Male	Dark Fitz + 65+ + Other	Light Fitz + 19-45 + Female	Light Fitz + 19-45 + Male	Light Fitz + 19-45 + Other	Light Fitz + 45-64 + Female	Light Fitz + 45-64 + Male	Light Fitz + 45-64 + Other	Light Fitz + 65+ + Female
Dark Fitz + 19-45 + Male	0															
Dark Fitz + 19-45 + Other	0.675	0														
Dark Fitz + 45-64 + Female	0.767	0	0.657													
Dark Fitz + 45-64 + Male	0	0.00002	0	0												
Dark Fitz + 45-64 + Other	0.009	0	0.047	0.011	0											
Dark Fitz + 65+ + Female	0.00000	0.020	0.00005	0.00000	0.00000	0.00000										
Dark Fitz + 65+ + Male	0	0.001	0	0	0.389	0	0.00001									
Dark Fitz + 65+ + Other	0	0	0	0	0	0	0	0								
Light Fitz + 19-45 + Female	0	0	0	0	0	0	0	0	0							
Light Fitz + 19-45 + Male	0	0	0	0	0	0.00000	0	0	0	0						
Light Fitz + 19-45 + Other	0.0001	0	0.006	0.0002	0	0.562	0	0	0	0	0.00003					
Light Fitz + 45-64 + Female	0.002	0	0.130	0.001	0	0.336	0	0	0	0	0	0.044				
Light Fitz + 45-64 + Male	0	0.089	0.00000	0	0.00000	0	0.200	0.00002	0	0	0	0	0			
Light Fitz + 45-64 + Other	0	0	0	0	0	0	0	0	0	0.063	0.00000	0	0	0		
Light Fitz + 65+ + Female	0.00001	0.0002	0.001	0.00002	0	0.00000	0.253	0.00000	0	0	0	0	0	0.020	0	
Light Fitz + 65+ + Male	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Age and Gender on Yolov5 and CCD can be found in Table 100

Table 100: AP. Pairwise Wilcox test with Bonferroni correction for Skin Type and the interaction with Age and Gender on Yolov5 and CCD

	Dark Fitz + 19-45 + Female	Dark Fitz + 19-45 + Male	Dark Fitz + 19-45 + Other	Dark Fitz + 45-64 + Female	Dark Fitz + 45-64 + Male	Dark Fitz + 45-64 + Other	Dark Fitz + 65+ + Female	Dark Fitz + 65+ + Male	Dark Fitz + 65+ + Other	Light Fitz + 19-45 + Female	Light Fitz + 19-45 + Male	Light Fitz + 19-45 + Other	Light Fitz + 45-64 + Female	Light Fitz + 45-64 + Male	Light Fitz + 45-64 + Other	Light Fitz + 65+ + Female
Dark Fitz + 19-45 + Male	0															
Dark Fitz + 19-45 + Other	0	0														
Dark Fitz + 45-64 + Female	0.861	0	0													
Dark Fitz + 45-64 + Male	0	0.933	0	0												
Dark Fitz + 45-64 + Other	0.029	0.198	0.00000	0.026	0.224											
Dark Fitz + 65+ + Female	0	0.001	0	0	0.003	0.003										
Dark Fitz + 65+ + Male	0	0	0	0	0	0	0									
Dark Fitz + 65+ + Other	0	0.00000	0	0	0.00000	0.00000	0.00000	0.142								
Light Fitz + 19-45 + Female	0	0	0.490	0	0	0	0	0	0							
Light Fitz + 19-45 + Male	0	0	0.753	0	0	0	0	0	0	0.502						
Light Fitz + 19-45 + Other	0	0	0.014	0	0	0	0	0	0	0.026	0.008					
Light Fitz + 45-64 + Female	0.091	0	0	0.006	0.00000	0.224	0	0	0	0	0	0				
Light Fitz + 45-64 + Male	0	0.050	0	0	0.067	0.425	0.00000	0	0.00000	0	0	0	0.0003			
Light Fitz + 45-64 + Other	0.00004	0	0.093	0.00004	0	0.00000	0	0	0	0.150	0.092	0.708	0.00000	0		
Light Fitz + 65+ + Female	0	0.634	0	0	0.654	0.241	0.002	0	0.00001	0	0	0	0.0005	0.408	0	
Light Fitz + 65+ + Male	0	0	0	0	0	0.00000	0.001	0	0.0001	0	0	0	0	0	0	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Lighting on AWS and CCD can be found in Table 101

Table 101: AP. Pairwise Wilcox test with Bonferroni correction for Lighting on AWS and CCD

	Bright
Dim	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Lighting on Azure and CCD can be found in Table 102

Table 102: AP. Pairwise Wilcox test with Bonferroni correction for Lighting on Azure and CCD

	Bright
Dim	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Lighting on GCP and CCD can be found in Table 103

Table 103: AP. Pairwise Wilcox test with Bonferroni correction for Lighting on GCP and CCD

	Bright
Dim	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Lighting on MogFace and CCD can be found in Table 104

Table 104: AP. Pairwise Wilcox test with Bonferroni correction for Lighting on MogFace and CCD

	Bright
Dim	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Lighting on TinaFace and CCD can be found in Table 105

Table 105: AP. Pairwise Wilcox test with Bonferroni correction for Lighting on TinaFace and CCD

	Bright
Dim	0

AP p-values for pairwise Wilcox test with Bonferroni correction for Lighting on Yolov5 and CCD can be found in Table 106

Table 106: AP. Pairwise Wilcox test with Bonferroni correction for Lighting on Yolov5 and CCD

	Bright
Dim	0

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] See Section 3.
   (b) Did you mention the license of the assets? [Yes] See Section 3.
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] See Section 3, and references within each of the papers that introduce the datasets that we use and the noise models that we use.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] See Section 3.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]