Supplementary Materials
Beyond Accuracy: A Framework for Evaluating Algorithmic Bias and Performance, Applied to Automated Sleep Scoring

Bechny Michal et al. (2025)

bechnymichal@gmail.com

1 Statistical characteristics of derived PSG markers

Metric	Sleep Scoring	Mean	SD	Q10	Q25	Q50	Q75	Q90	Min	Max
Sleep Latency [*minutes*]	-	17.8	23.9	2.0	4.5	10.5	21.5	40.5	0.0	339.5
	U-Sleep	15.4	20.8	1.5	4.0	8.5	18.0	36.0	0.0	204.5
	YASA	26.3	29.2	4.0	8.5	17.0	33.0	60.0	0.0	306.5
REM Latency [*minutes*]	-	139.5	80.9	59.0	77.5	119.0	184.5	258.9	0.0	502.0
	U-Sleep	131.1	84.7	48.0	72.5	111.0	177.0	252.0	0.0	896.5
	YASA	112.4	66.4	45.2	67.5	97.5	149.5	201.5	0.0	486.5
Total Sleep Time [*minutes*]	-	338.3	89.0	239.5	291.5	337.5	382.0	424.8	0.0	848.5
	U-Sleep	343.6	89.4	247.0	297.0	342.0	385.0	428.5	0.0	865.0
	YASA	301.1	94.0	192.2	248.8	302.5	351.0	398.0	0.0	768.5
WASO [*minutes*]	-	64.2	53.8	12.5	25.0	50.0	89.2	135.0	0.0	952.0
	U-Sleep	61.3	52.5	12.5	23.5	46.5	84.0	132.0	0.0	932.0
	YASA	92.8	62.8	29.0	46.5	78.0	124.5	174.0	3.0	980.0
Sleep Cycles [*N*]	-	2.6	1.4	1.0	1.5	2.5	3.5	4.0	0.0	11.0
	U-Sleep	2.8	1.4	1.0	2.0	2.5	3.5	4.5	0.0	10.5
	YASA	2.6	1.2	1.0	1.5	2.5	3.5	4.0	0.0	9.5
Sleep-stage Transitions [$N/hour$]	-	21.2	7.3	12.6	16.1	20.5	25.3	30.8	1.2	64.4
	U-Sleep	14.6	4.9	9.3	11.3	13.9	17.2	20.8	0.9	62.3
	YASA	17.1	5.4	11.2	13.5	16.4	19.9	23.9	0.2	69.8
Awakenings [$N/hour$]	-	3.5	2.2	1.4	2.1	3.1	4.4	6.0	0.0	27.9
	U-Sleep	3.3	1.8	1.5	2.2	3.0	4.2	5.5	0.0	20.6
	YASA	4.9	2.3	2.6	3.3	4.5	6.0	7.9	0.2	28.6
Sleep Efficiency [%]	-	80.1	14.9	60.4	73.4	83.7	91.2	94.9	0.0	100.0
	U-Sleep	81.5	14.4	62.4	75.4	85.3	91.8	95.2	0.0	100.0
	YASA	71.3	17.2	48.4	62.5	74.9	84.5	89.4	0.0	98.1
W [%]	-	19.9	14.9	5.1	8.8	16.3	26.6	39.6	0.0	100.0
	U-Sleep	18.5	14.4	4.8	8.2	14.7	24.6	37.6	0.0	100.0
	YASA	28.7	17.2	10.6	15.5	25.1	37.5	51.6	1.9	100.0
N1 [%]	-	15.9	10.3	5.8	8.7	13.4	20.3	29.1	0.0	85.5
	U-Sleep	10.6	7.8	3.5	5.4	8.5	13.4	20.0	0.0	69.2
	YASA	5.1	3.8	1.2	2.6	4.4	6.8	9.5	0.0	39.2
N2 [%]	-	35.5	12.3	19.2	27.7	36.3	43.9	50.1	0.0	87.4
	U-Sleep	44.5	12.1	29.1	37.8	45.5	52.1	58.5	0.0	96.9
	YASA	39.9	10.1	27.2	34.6	41.0	46.4	51.4	0.0	73.5
N3 [%]	-	16.2	10.3	2.7	9.0	15.4	22.3	29.3	0.0	75.3
	U-Sleep	13.1	8.6	1.4	6.7	12.8	18.7	24.0	0.0	58.6
	YASA	13.7	8.4	1.6	7.3	13.7	19.6	24.7	0.0	48.6
REM [%]	-	12.6	6.9	3.3	7.8	12.5	17.2	21.4	0.0	48.5
	U-Sleep	13.3	7.2	3.4	8.3	13.4	18.1	22.3	0.0	54.9
	YASA	12.6	6.5	3.5	8.0	12.7	17.3	21.0	0.0	44.1

Table 1. The descriptive statistics for sleep metrics from the reference dataset are presented, along with predictions based on two models: U-Sleep and YASA. For each metric, the mean, standard deviation (SD), quantiles (Q10, Q25, Q50, Q75, Q90), minimum (Min), and maximum (Max) are reported. The paired Wilcoxon signed-rank test was used to compare model predictions against physicians-based reference values, testing the null hypothesis that the differences between them are symmetrically distributed around zero. Significant results highlight the corresponding model's name (U-Sleep or YASA) based on the test p-value thresholds: 0.05 , 0.01 , and 0.001 , respectively.

2 Statistical characteristics of raw errors in algorithm-derived PSG markers

Metric-Error	Algorithm	Mean	SD	Q10	Q25	Q50	Q75	Q90	Min	Max
Sleep Latency	U-Sleep	-2.5	14.7	-10.0	-2.0	0.0	0.5	3.0	-271.0	160.5
[*minutes*]	YASA	8.5	19.2	0.0	1.0	3.5	10.0	22.5	-266.5	262.5
REM Latency	U-Sleep	-0.3	69.1	-45.6	-2.5	0.0	3.5	44.5	-410.5	608.0
[*minutes*]	YASA	-18.4	79.0	-112.5	-28.0	-4.5	0.0	50.8	-394.5	411.5
Total Sleep Time	U-Sleep	5.2	24.6	-12.0	-3.5	2.0	10.5	25.5	-353.5	263.0
[*minutes*]	YASA	-37.3	42.3	-81.5	-47.0	-26.0	-14.0	-6.0	-532.5	164.5
WASO	U-Sleep	-2.8	24.1	-21.5	-8.5	-1.5	4.0	14.5	-259.0	317.5
[*minutes*]	YASA	28.8	38.5	1.0	9.0	20.0	38.8	68.0	-242.0	426.5
Sleep Cycles	U-Sleep	0.2	0.7	0.0	0.0	0.0	0.0	1.0	-8.0	5.0
[*N*]	YASA	-0.0	0.9	-1.0	0.0	0.0	0.0	1.0	-10.5	3.5
Sleep-stage Transitions	U-Sleep	-6.5	6.0	-13.7	-9.9	-6.2	-3.0	-0.1	-39.8	36.0
[$N/hour$]	YASA	-4.0	6.4	-11.5	-7.8	-3.9	-0.3	3.4	-38.2	37.5
Awakenings	U-Sleep	-0.1	1.5	-1.5	-0.7	-0.1	0.5	1.3	-17.1	9.5
[$N/hour$]	YASA	1.4	2.2	-0.8	0.2	1.2	2.5	4.2	-12.7	15.6
Sleep Efficiency	U-Sleep	1.3	5.7	-2.7	-0.8	0.6	2.6	6.2	-66.3	63.4
[%]	YASA	-8.8	9.8	-19.4	-11.3	-6.2	-3.3	-1.5	-99.9	40.2
W	U-Sleep	-1.3	5.7	-6.2	-2.6	-0.6	0.8	2.7	-63.4	66.3
[%]	YASA	8.8	9.8	1.5	3.3	6.2	11.3	19.4	-40.2	99.9
N1	U-Sleep	-5.4	7.6	-14.4	-8.6	-4.1	-1.0	1.5	-61.5	41.1
[%]	YASA	-10.8	10.0	-23.4	-15.0	-8.6	-4.4	-1.7	-77.4	34.6
N2	U-Sleep	9.0	9.1	0.0	3.2	7.5	13.1	20.0	-30.1	78.3
[%]	YASA	4.5	8.9	-5.0	-0.7	3.7	9.0	15.3	-78.1	51.6
N3	U-Sleep	-3.1	6.1	-10.2	-5.7	-2.2	0.0	2.8	-75.3	30.6
[%]	YASA	-2.5	6.0	-9.7	-4.8	-1.4	0.5	3.2	-75.3	23.0
REM	U-Sleep	0.7	3.1	-1.9	-0.3	0.5	1.9	3.8	-37.6	35.3
[%]	YASA	0.0	4.1	-4.2	-1.5	0.3	2.1	4.1	-39.8	25.0

Table 2. Summary of the prediction errors for sleep metrics, derived from hypnograms predicted by U-Sleep and YASA compared to those derived from physician-scored hypnograms. The table presents the mean error, standard deviation (SD), (10, 25, 50, 75, 90)%-quantiles (Q10, Q25, Q50, Q75, Q90), minimum (Min), and maximum (Max) values. The paired Wilcoxon signed-rank test was used to compare model predictions against physicians-based reference values, testing the null hypothesis that the differences between them are symmetrically distributed around zero. Significant results highlight corresponding model's name (U-Sleep or YASA) based on the test p-value thresholds: 0.05 , 0.01 , and 0.001 , respectively.

3 Partial effects of age on U-Sleep and YASA performance metrics

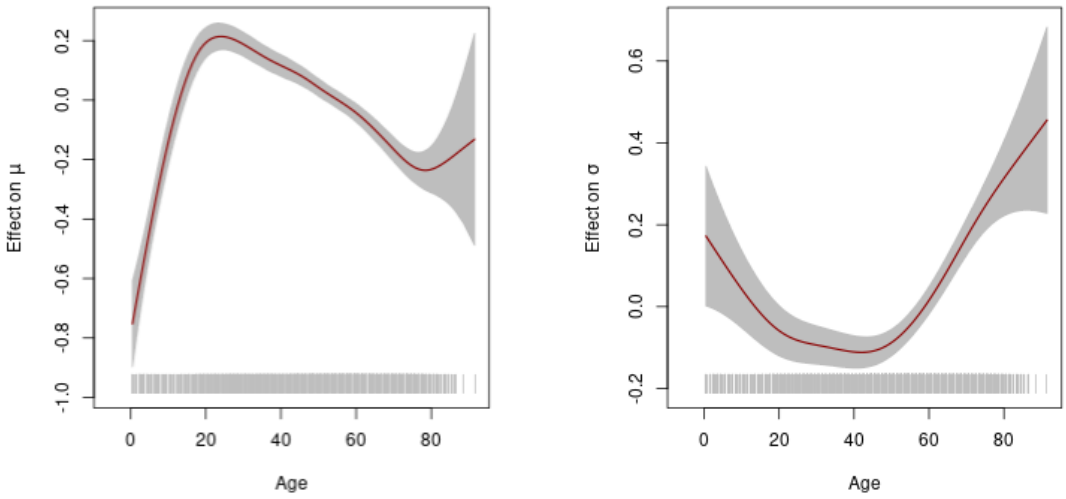

Figure 1. Partial effects of age on in U-Sleep accuracy. The left panel shows the estimated expected bias (location parameter) across ages, while the right panel illustrates its variability (scale parameter). Shaded areas represent 95% confidence intervals.

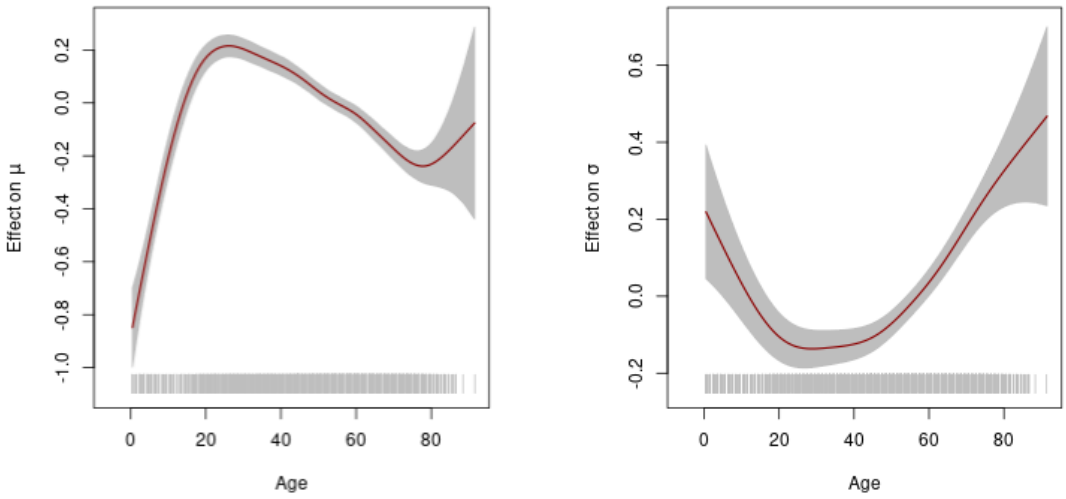

Figure 2. Partial effects of age on in U-Sleep F1-score. The left panel shows the estimated expected bias (location parameter) across ages, while the right panel illustrates its variability (scale parameter). Shaded areas represent 95% confidence intervals.

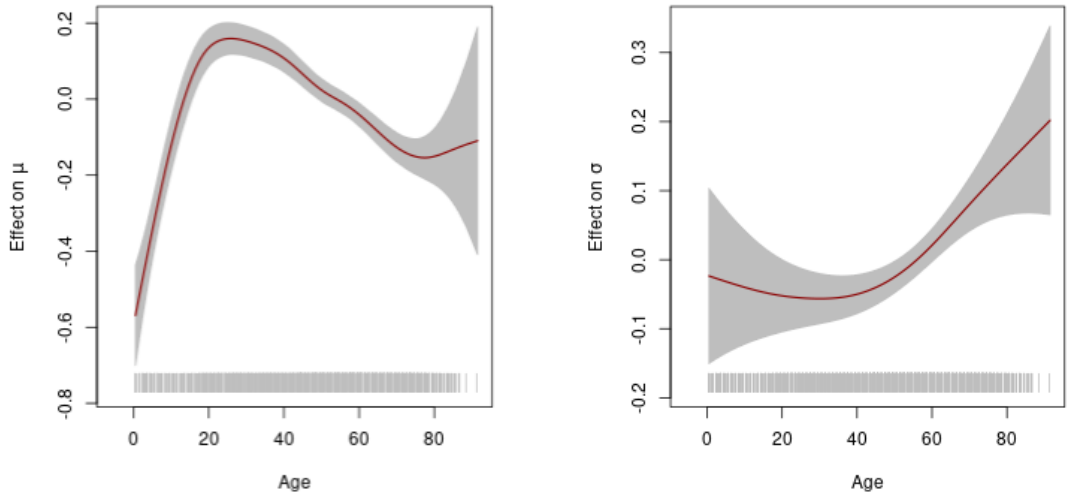

Figure 3. Partial effects of age on YASA accuracy. The left panel shows the estimated expected bias (location parameter) across ages, while the right panel illustrates its variability (scale parameter). Shaded areas represent 95% confidence intervals.

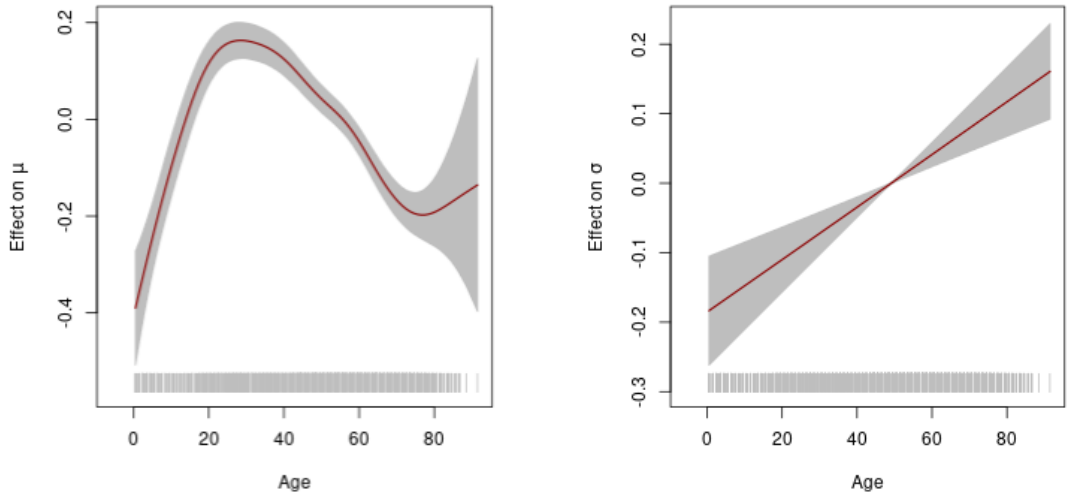

Figure 4. Partial effects of age on YASA F1-score. The left panel shows the estimated expected bias (location parameter) across ages, while the right panel illustrates its variability (scale parameter). Shaded areas represent 95% confidence intervals.

4 Performance Plots

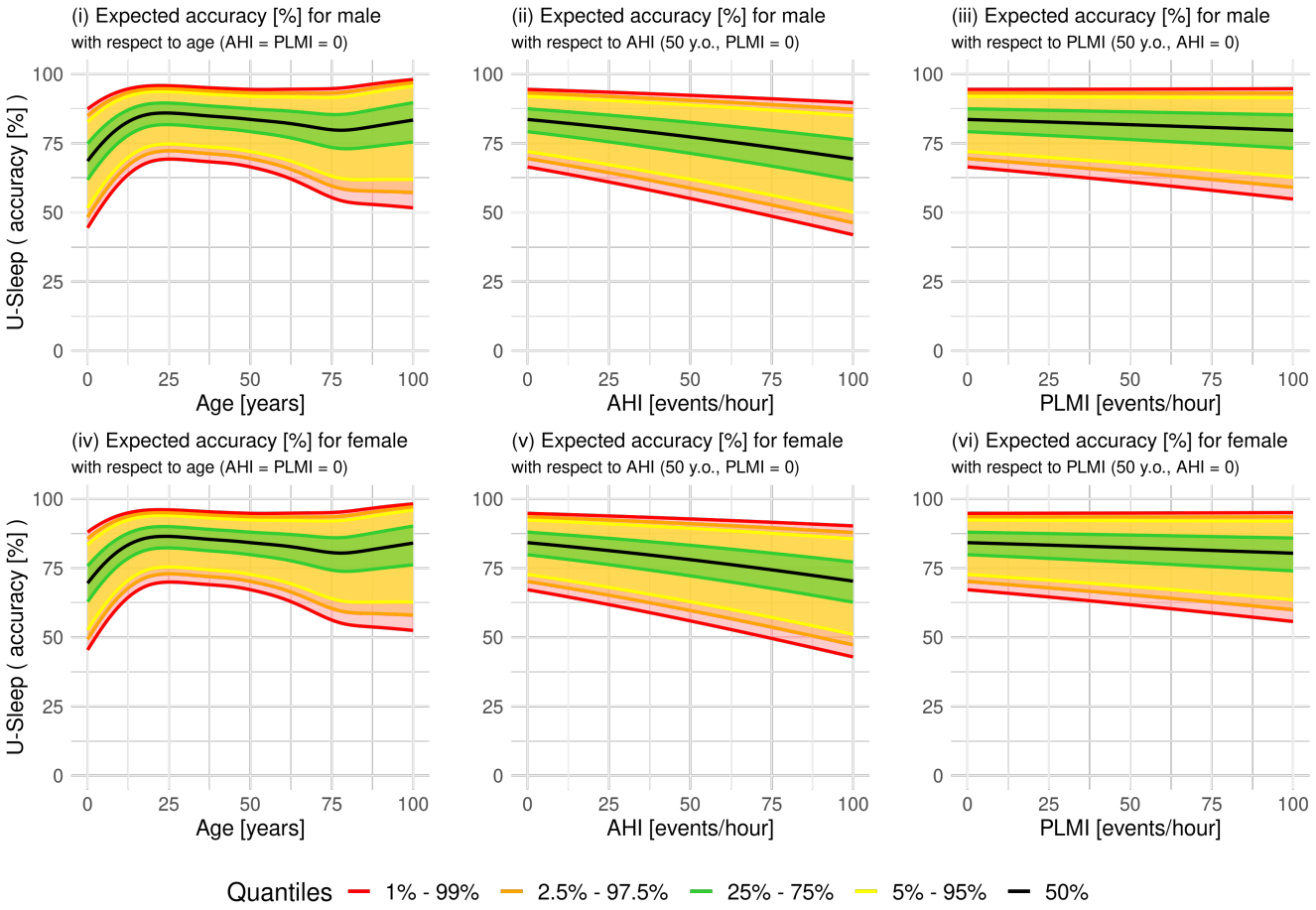

Figure 5. Expected distribution of the subject-specific accuracy based on the zero-and-ones-inflated Beta performance model for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, showing the expected performance variability across subjects' characteristics.

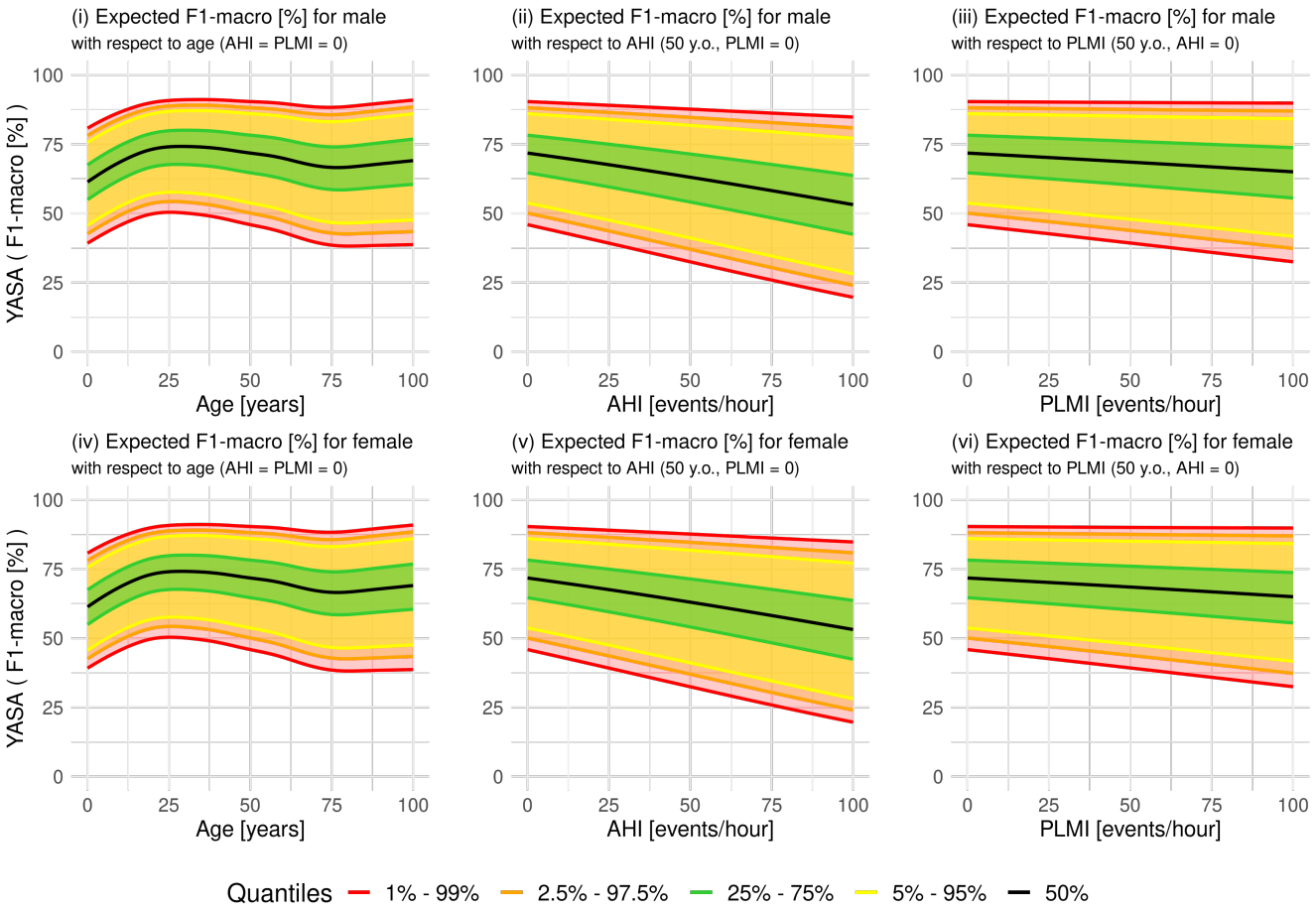

Figure 6. Expected distribution of the subject-specific macro F1-score based on the zero-and-ones-inflated Beta performance model for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, reflecting the expected performance variability across subjects' characteristics.

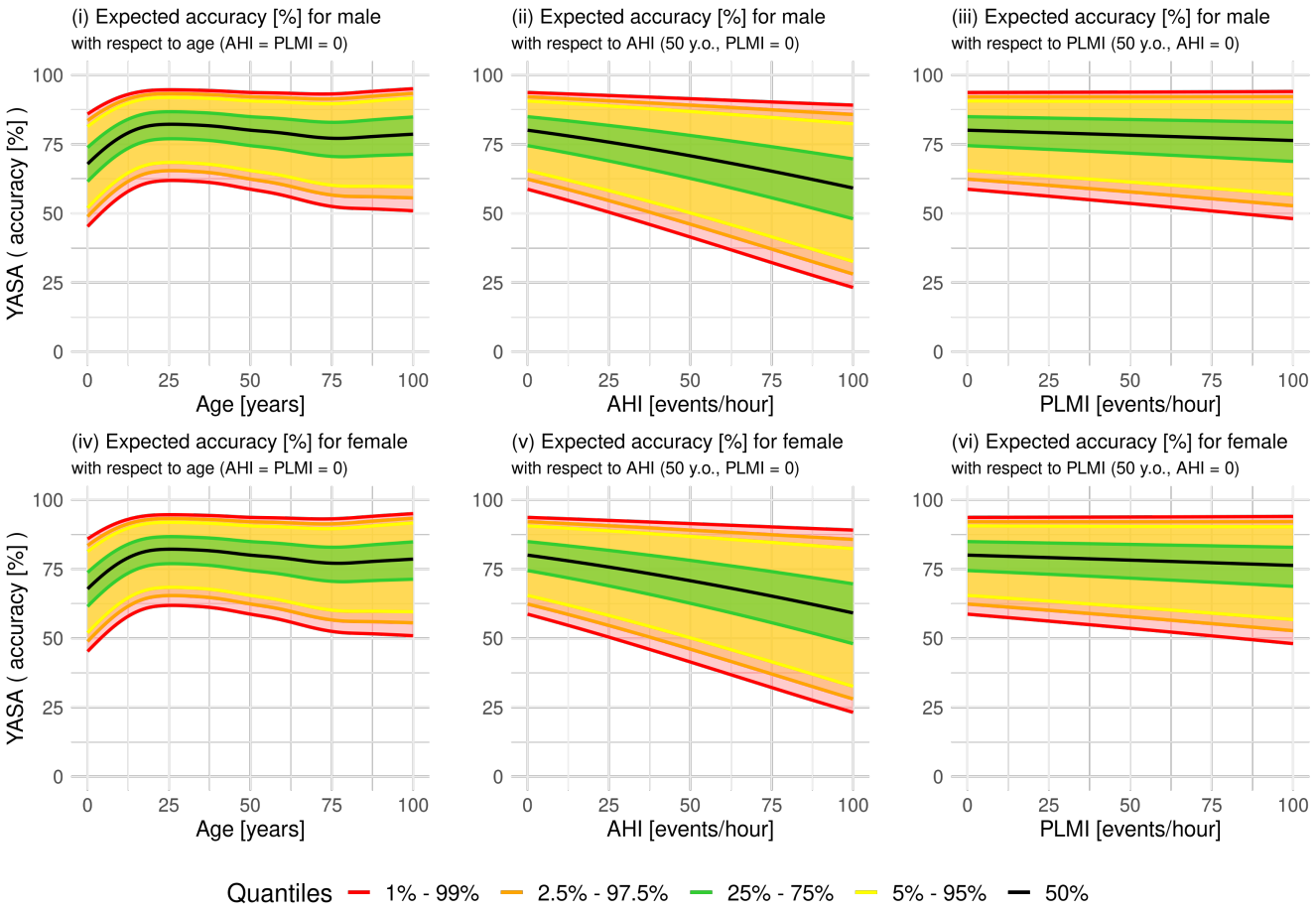

Figure 7. Expected distribution of the subject-specific accuracy based on the zero-and-ones-inflated Beta performance model for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, showing the expected performance variability across subjects' characteristics.

5 Partial effects of age on bias in U-Sleep and YASA derived percentage of wakefulness

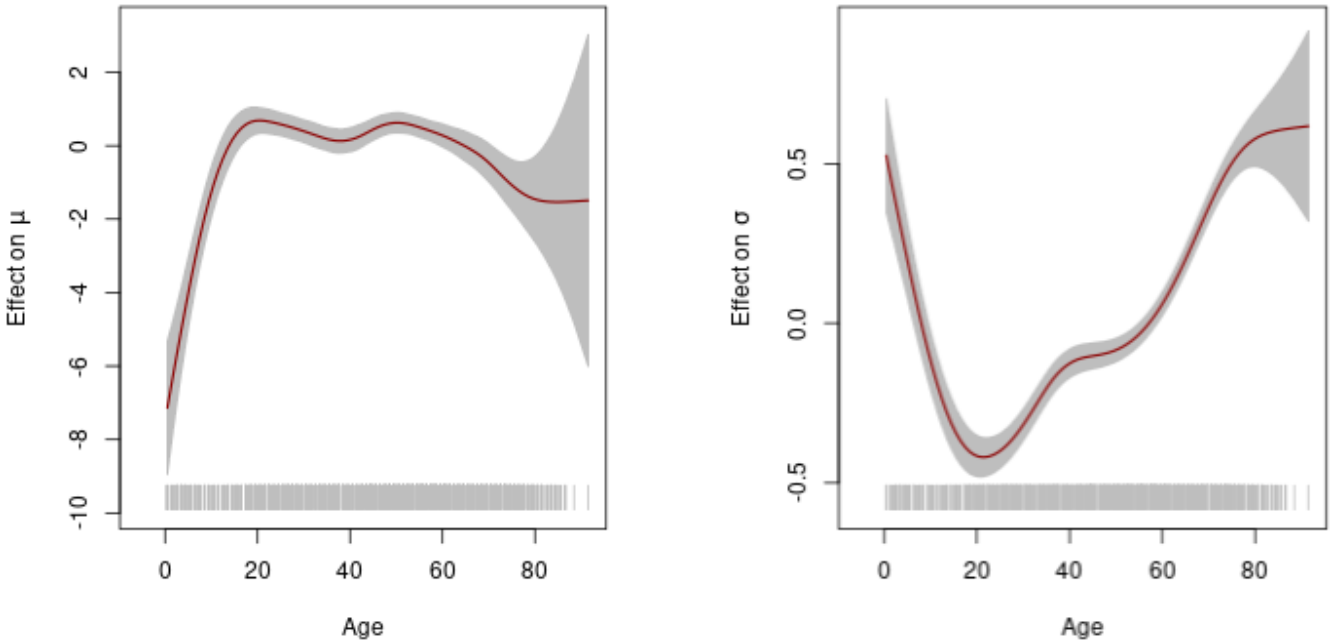

Figure 8. Partial effects of age on bias in U-Sleep-derived wakefulness percentage (W%). The left panel shows the estimated expected bias (location parameter) across ages, while the right panel illustrates its variability (scale parameter). Shaded areas represent 95% confidence intervals.

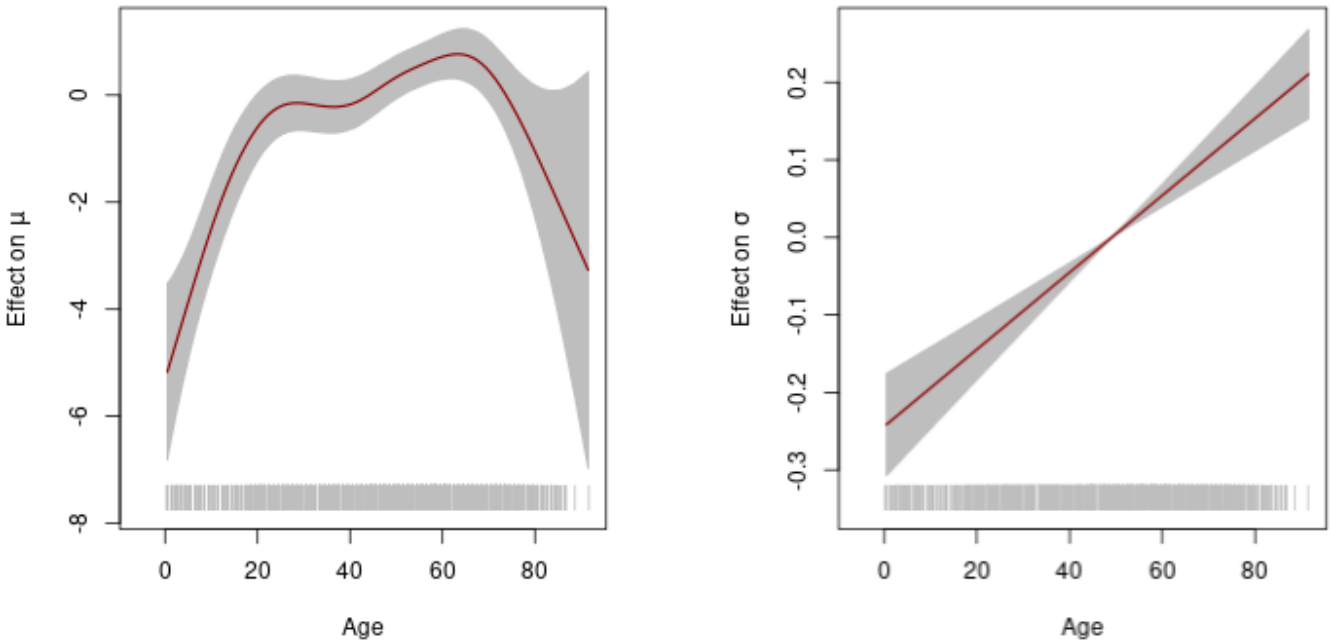

Figure 9. Partial effects of age on bias in YASA-derived wakefulness percentage (W%). The left panel shows the estimated expected bias (location parameter) across ages, while the right panel illustrates its variability (scale parameter). Shaded areas represent 95% confidence intervals.

6 Bias in clinical PSG markers based on YASA predictions

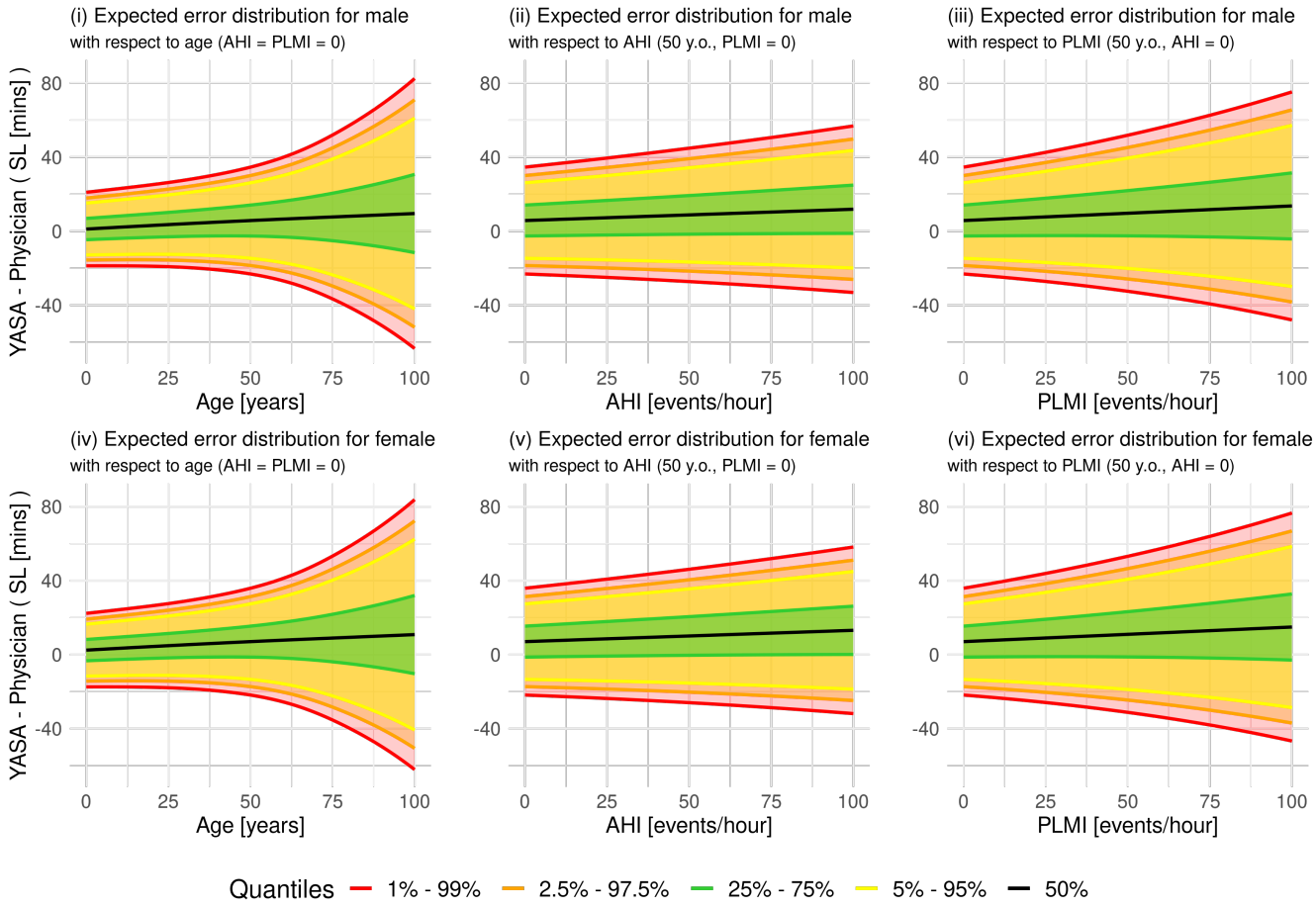

Figure 10. Expected distribution of the bias in the sleep latency (SL, minutes) based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

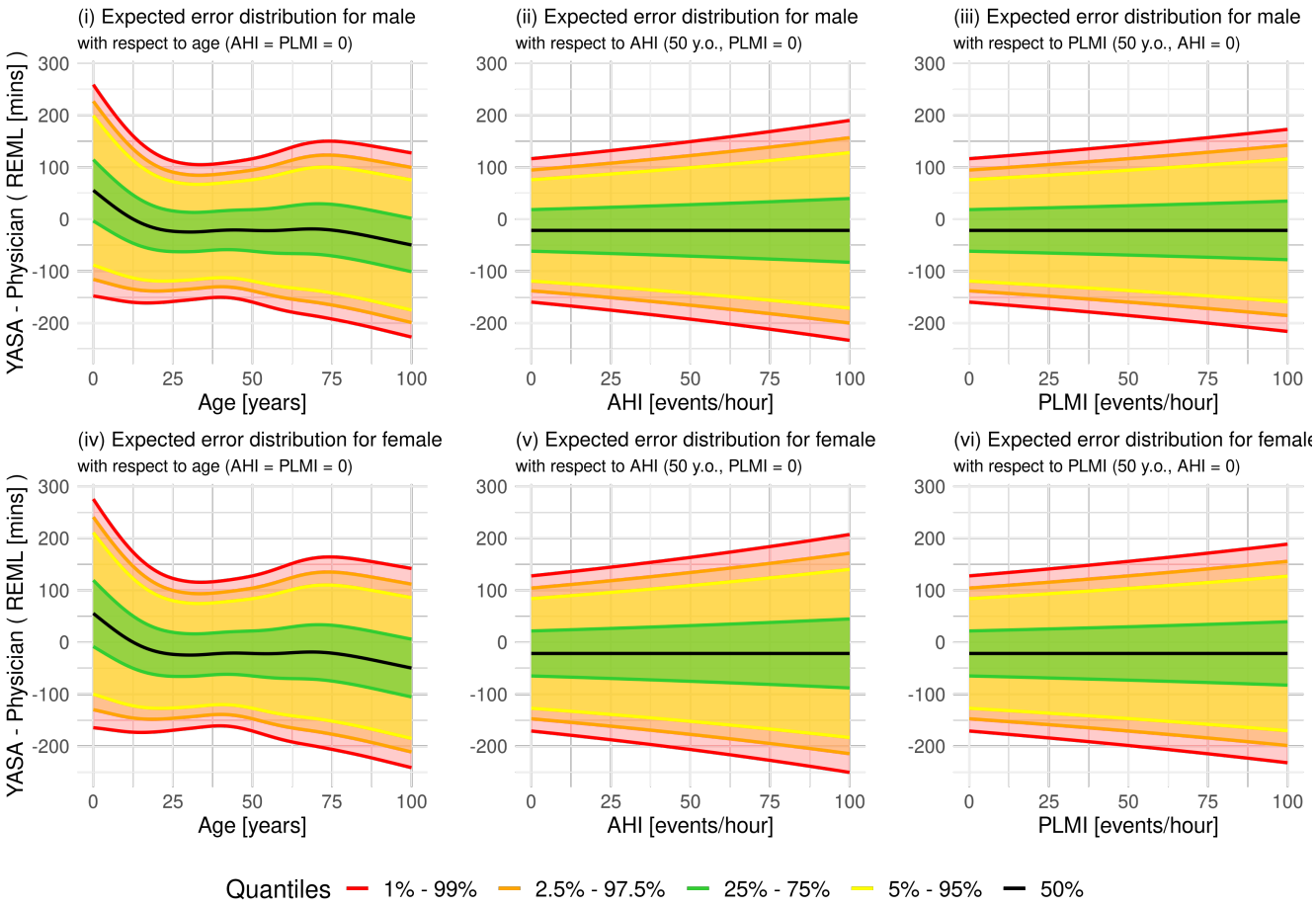

Figure 11. Expected distribution of the bias in the REM latency (REML, minutes) based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

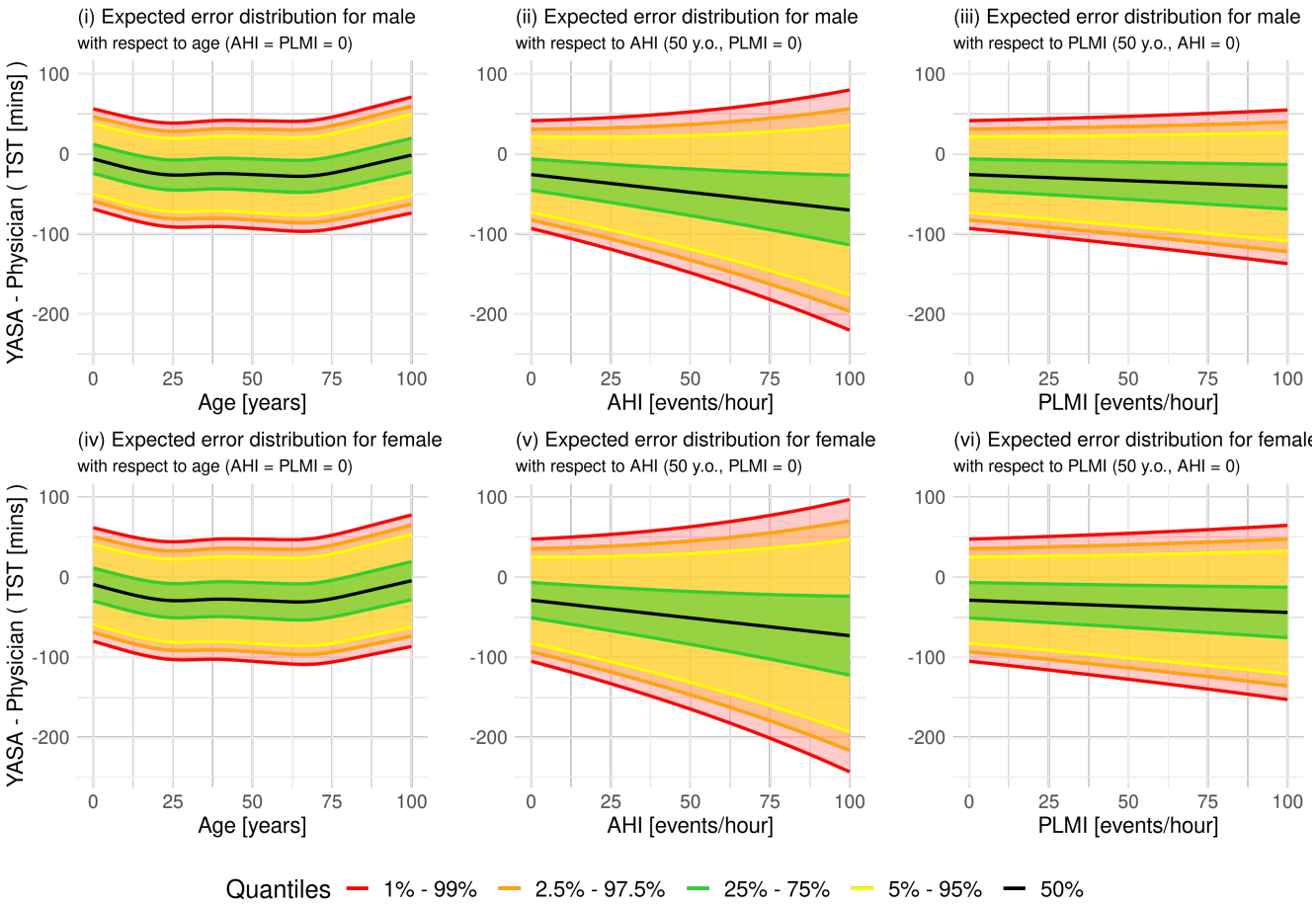

Figure 12. Expected distribution of the bias in the total sleep time (TST, minutes) based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

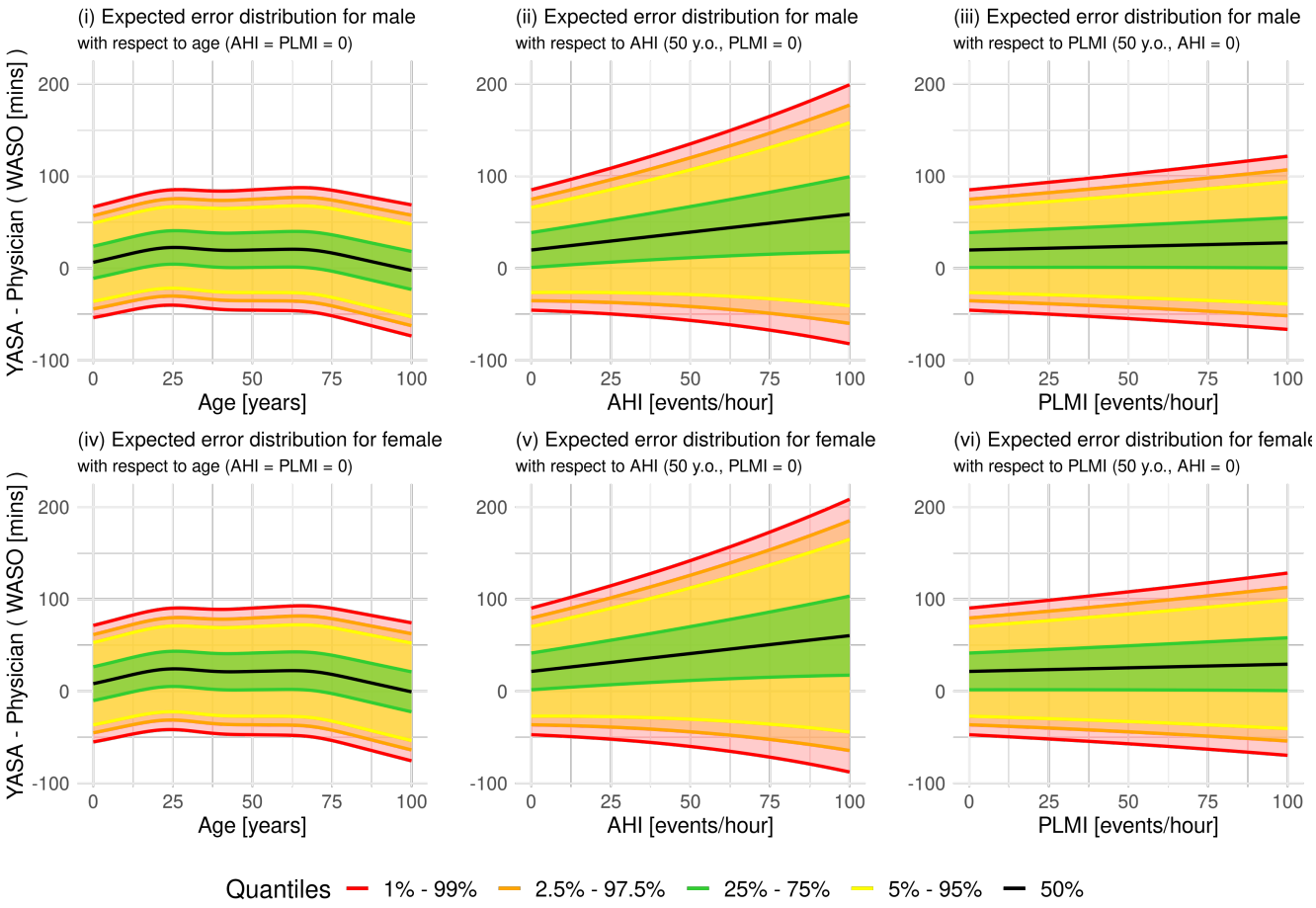

Figure 13. Expected distribution of the bias in the wake after sleep onset (WASO, minutes) based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

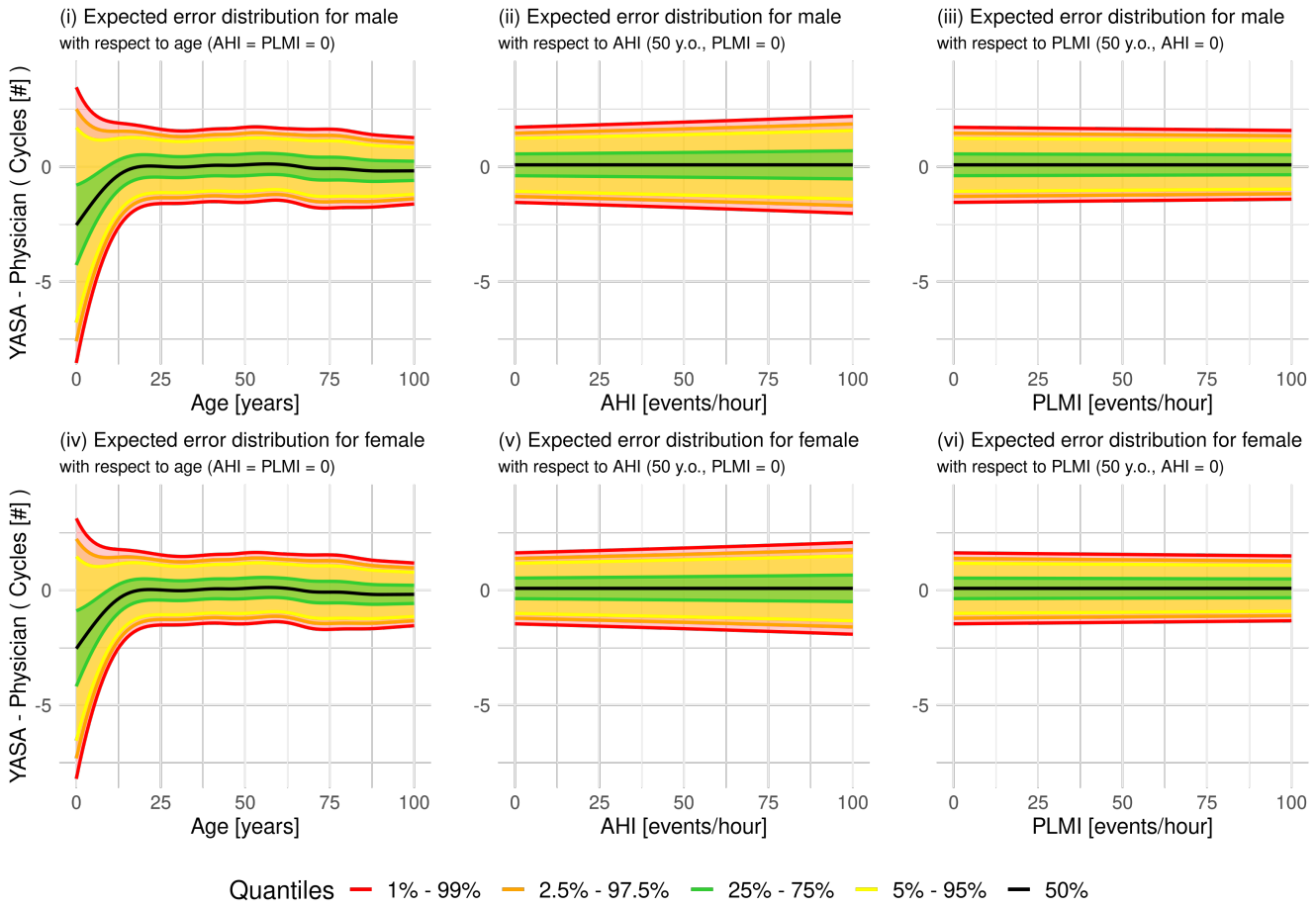

Figure 14. Expected distribution of the bias in the number (#) of sleep cycles based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

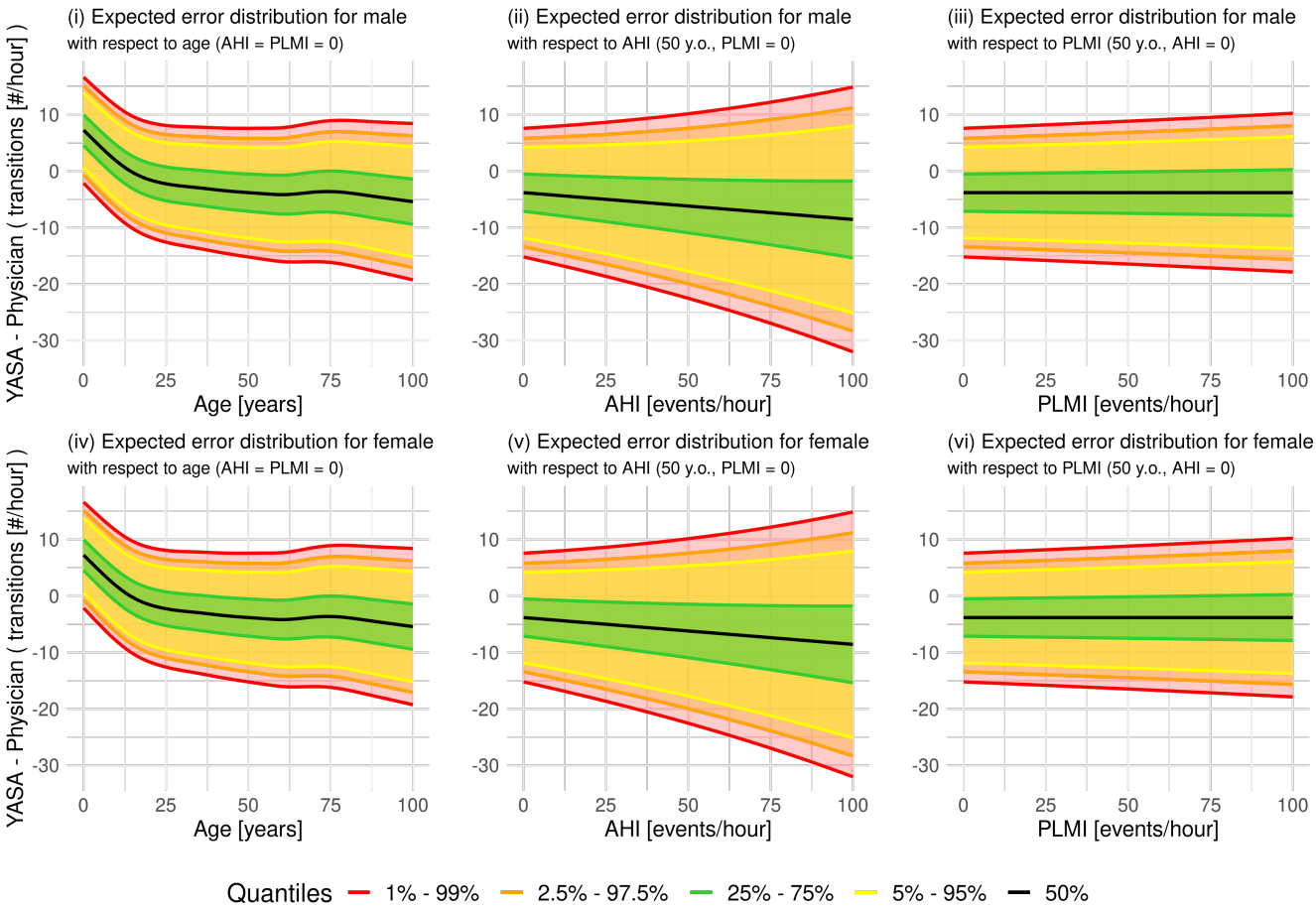

Figure 15. Expected distribution of the bias in the hourly rate (# / hour) of sleep stage transitions based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

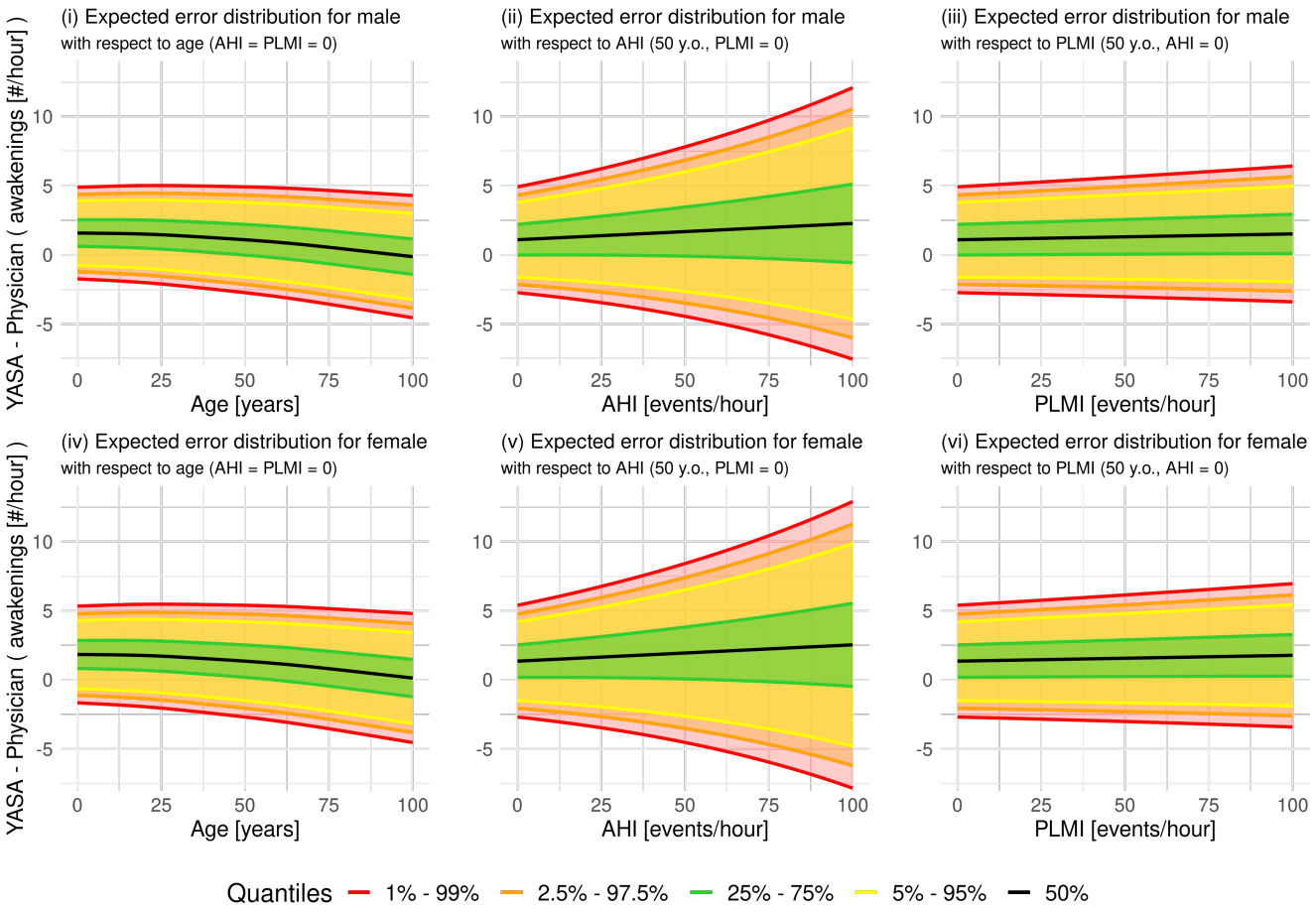

Figure 16. Expected distribution of the bias in the hourly rate (# / hour) of awakenings based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

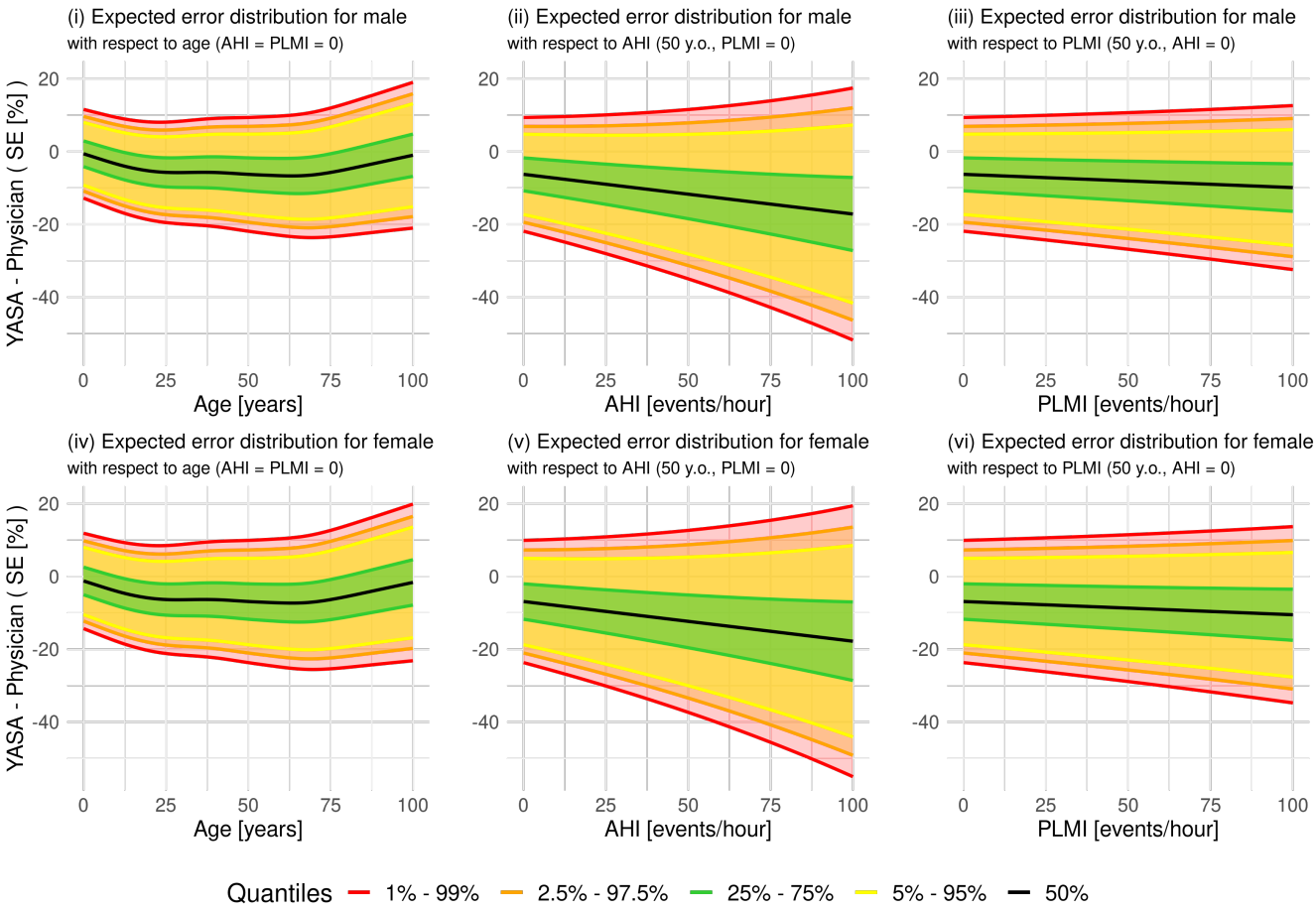

Figure 17. Expected distribution of the bias in the sleep efficiency percentage (SE, %) based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

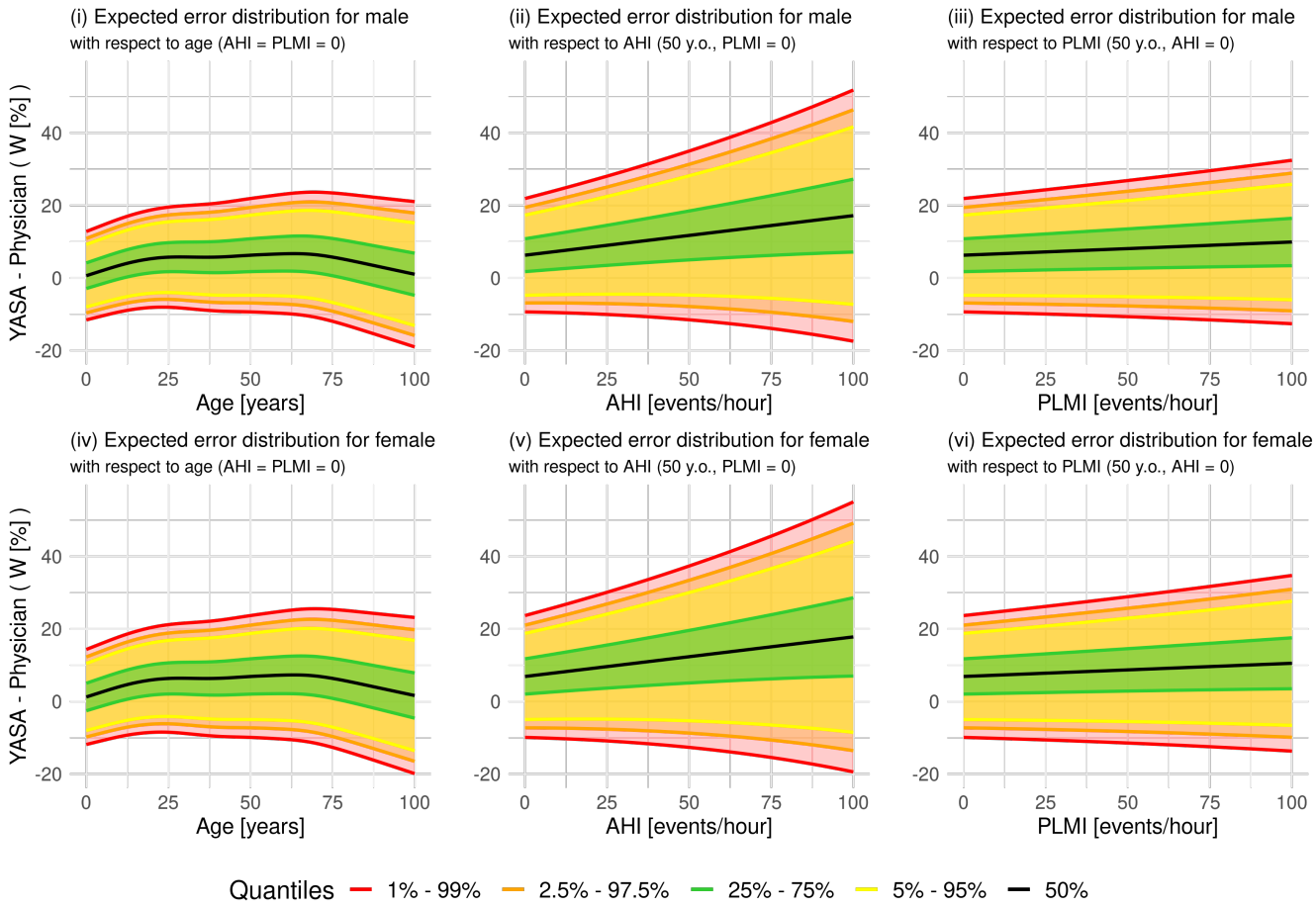

Figure 18. Expected distribution of the bias in the wakefulness percentage after sleep onset (W, %) based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

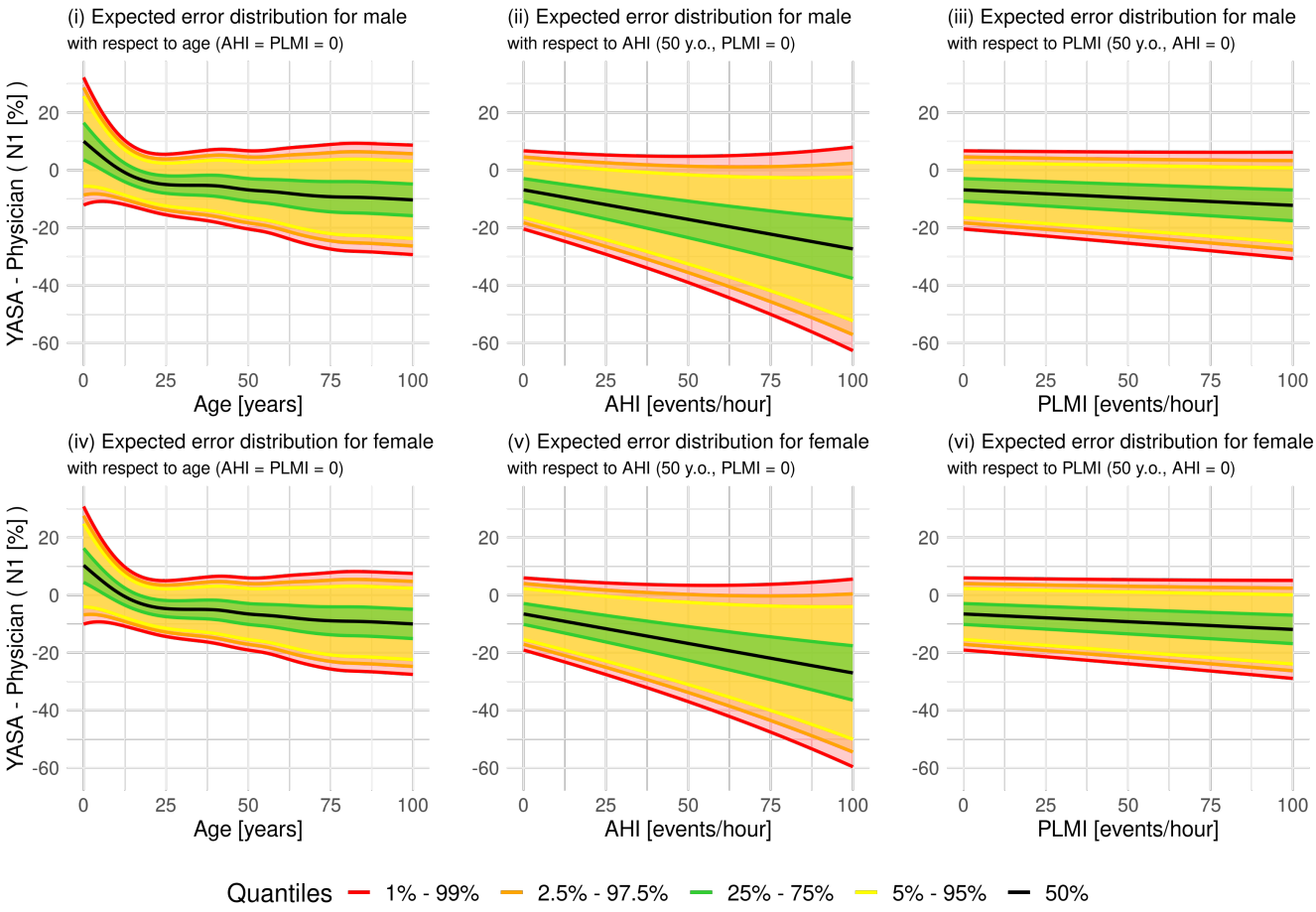

Figure 19. Expected distribution of the bias in the N1 sleep percentage after sleep onset (N1, %) based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

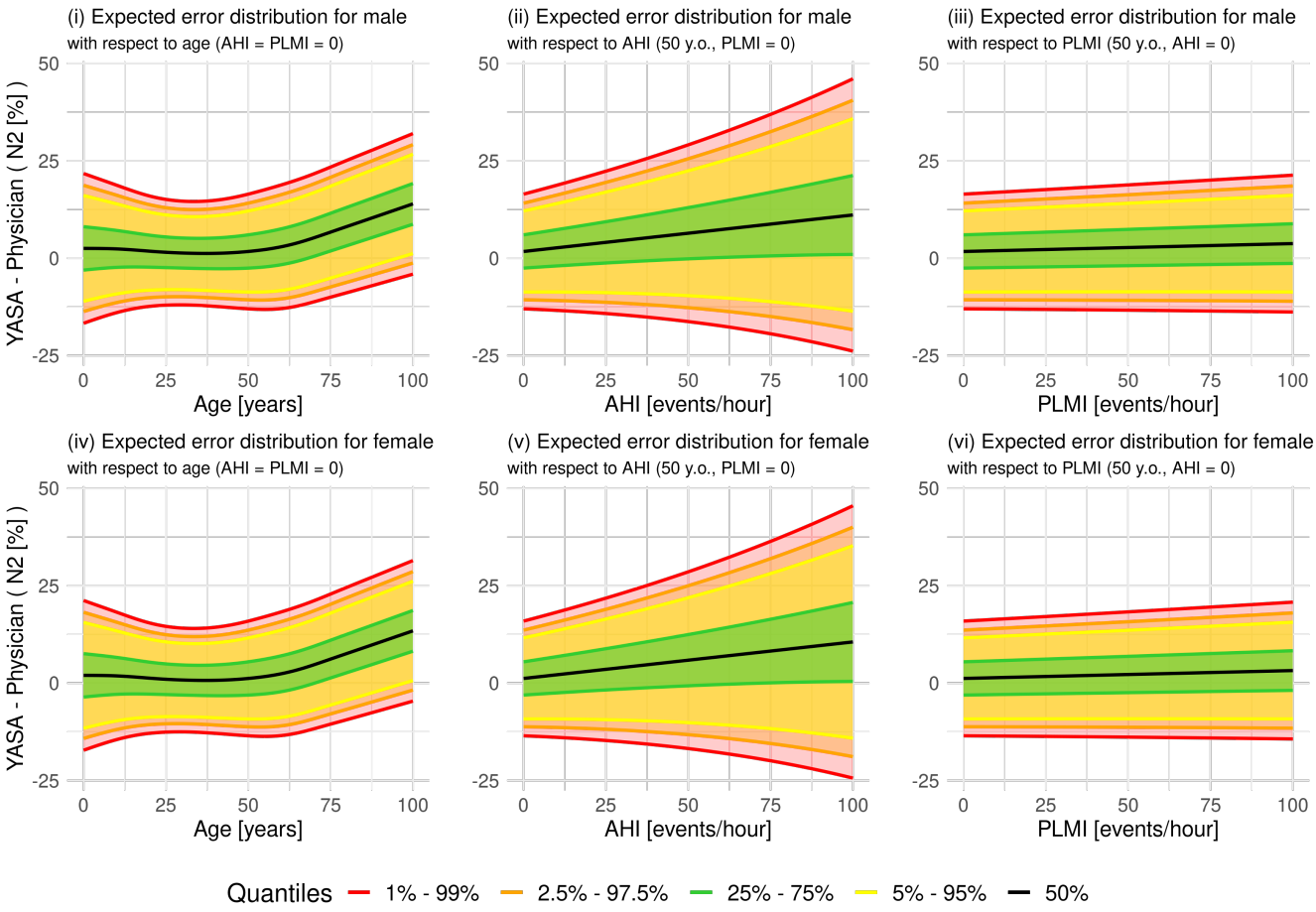

Figure 20. Expected distribution of the bias in the N2 sleep percentage after sleep onset (N2, %) based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

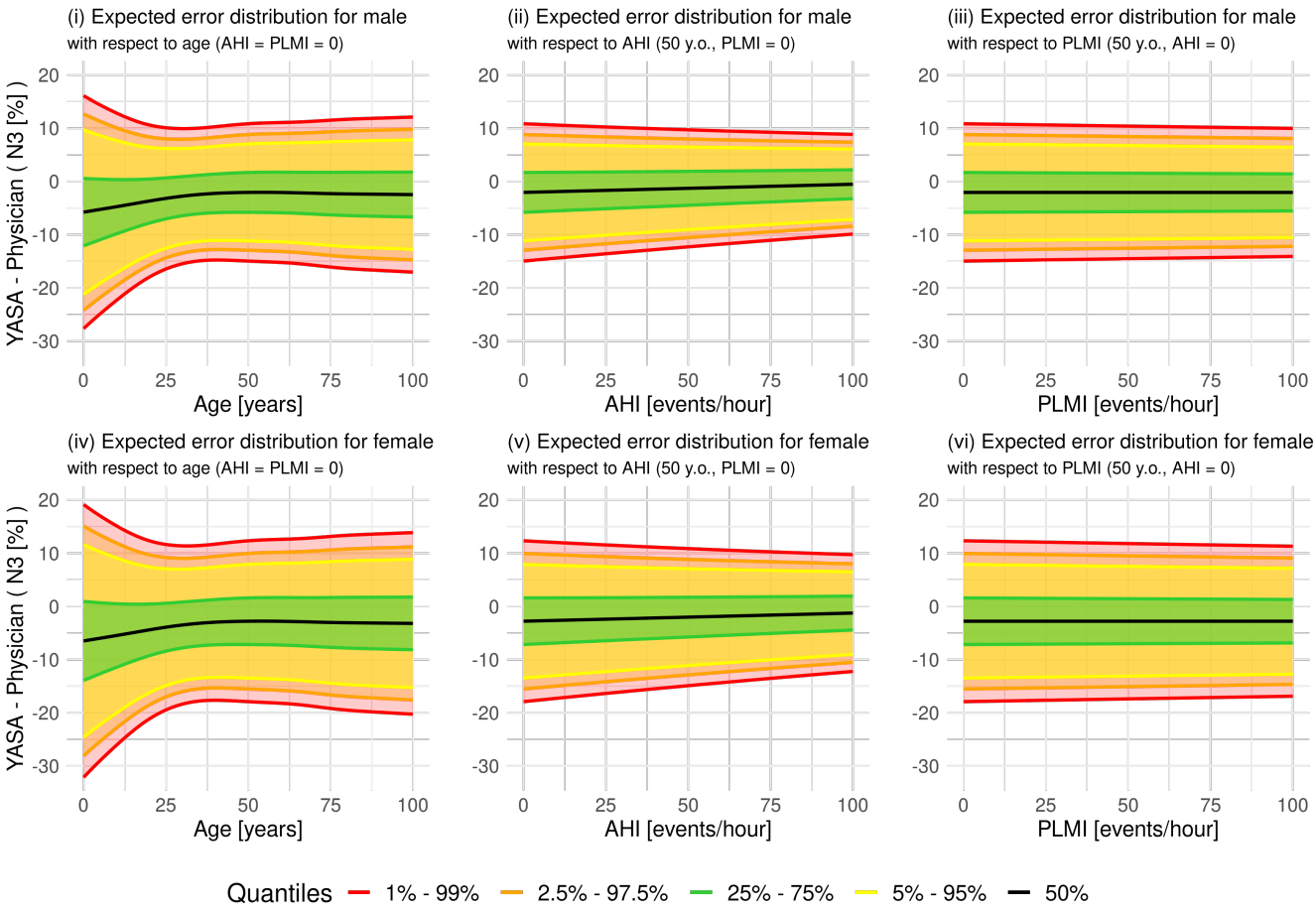

Figure 21. Expected distribution of the bias in the N3 sleep percentage after sleep onset (N3, %) based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

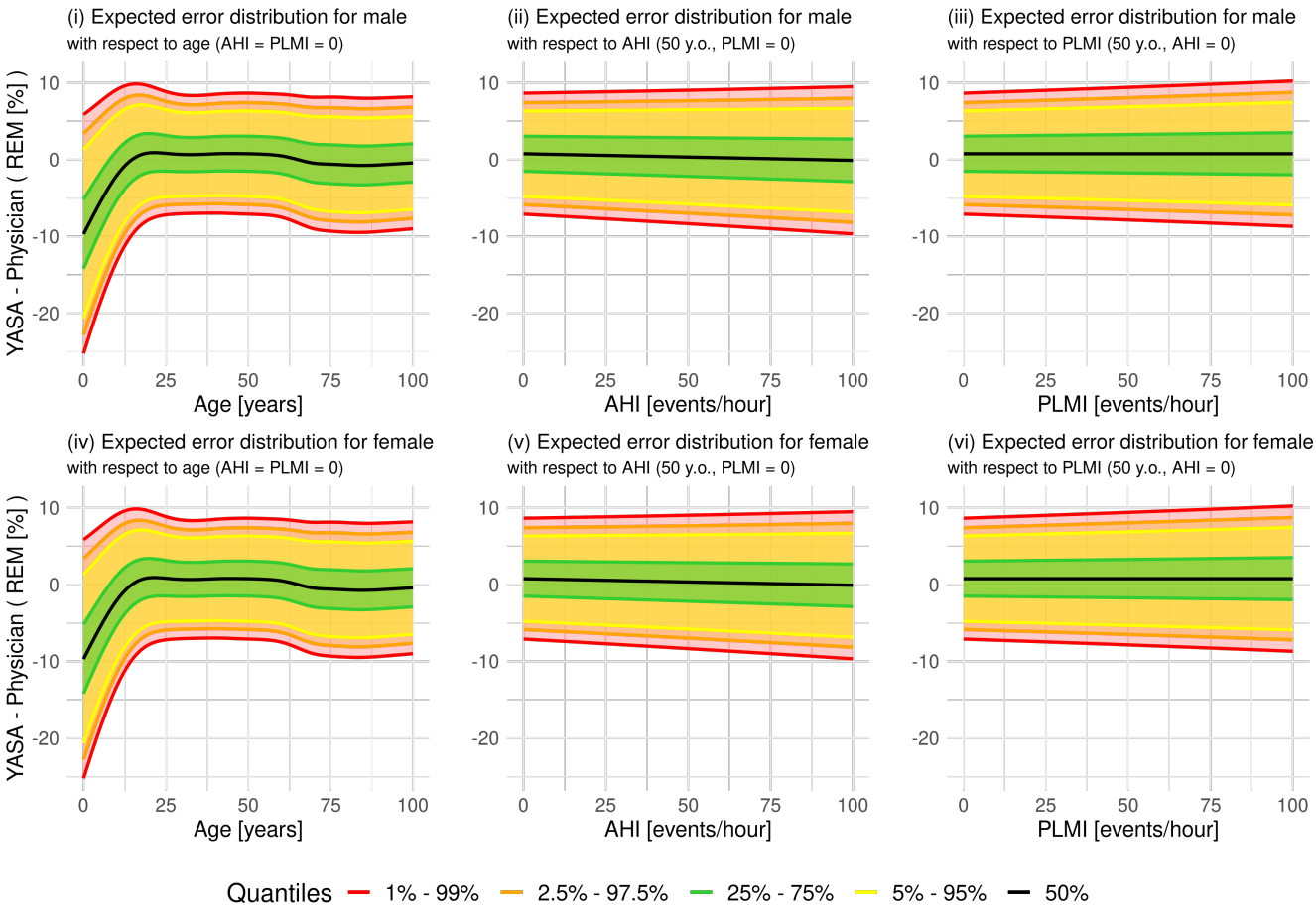

Figure 22. Expected distribution of the bias in the REM sleep percentage after sleep onset (REM, %) based on the generalized normal distribution for YASA predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

7 Bias in clinical PSG markers based on U-Sleep predictions

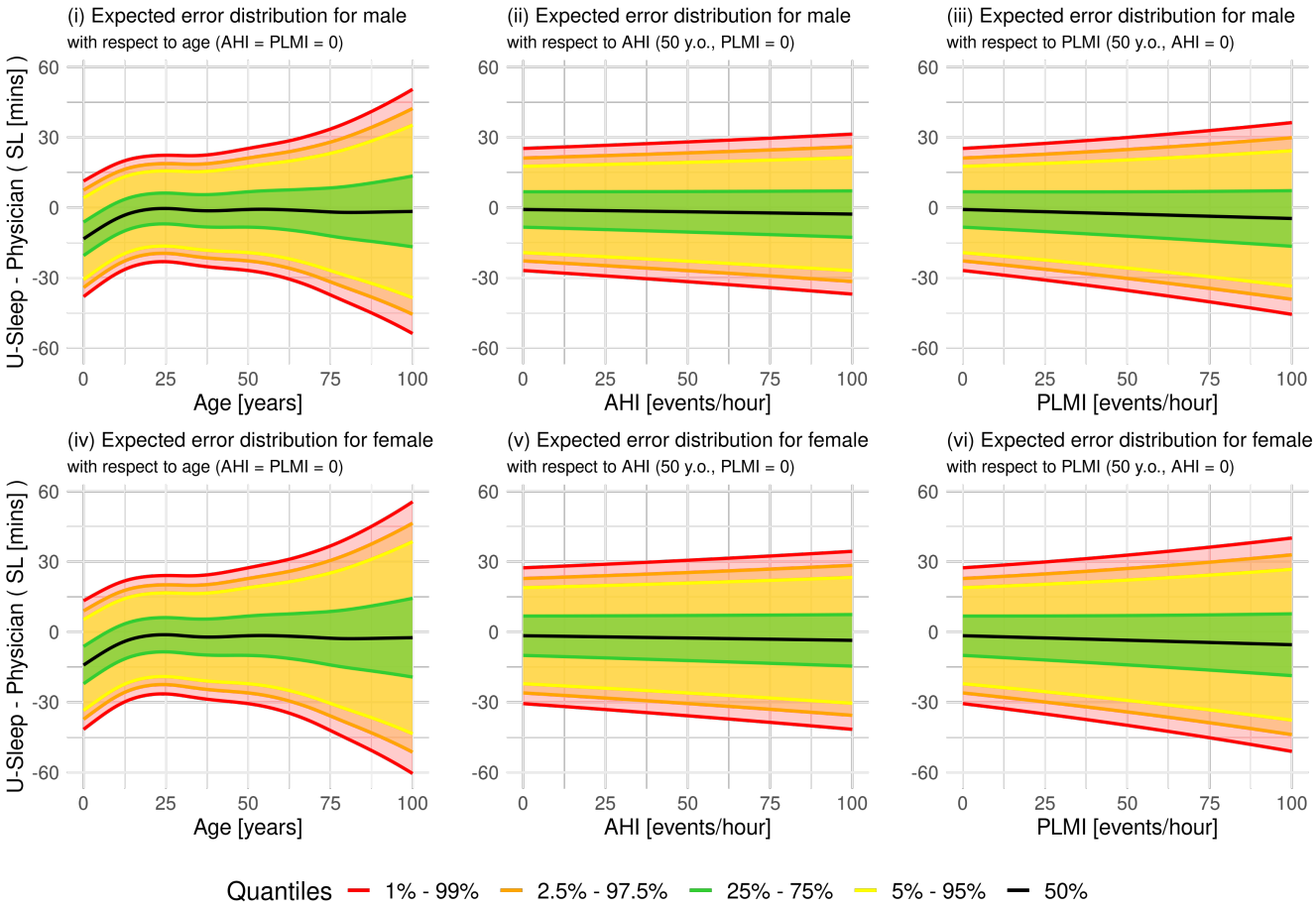

Figure 23. Expected distribution of the bias in the sleep latency (SL, minutes) based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

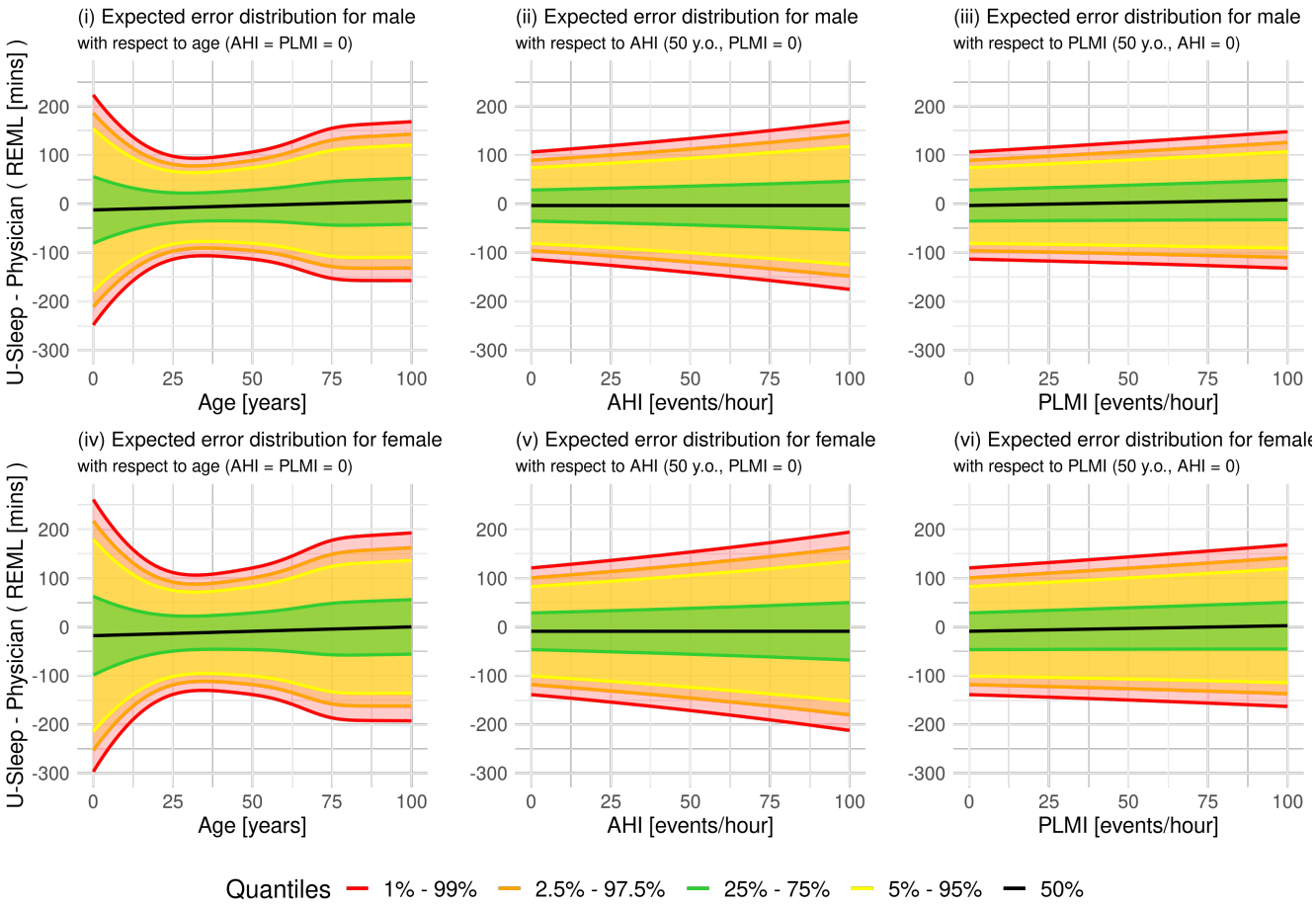

Figure 24. Expected distribution of the bias in the REM latency (REML, minutes) based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

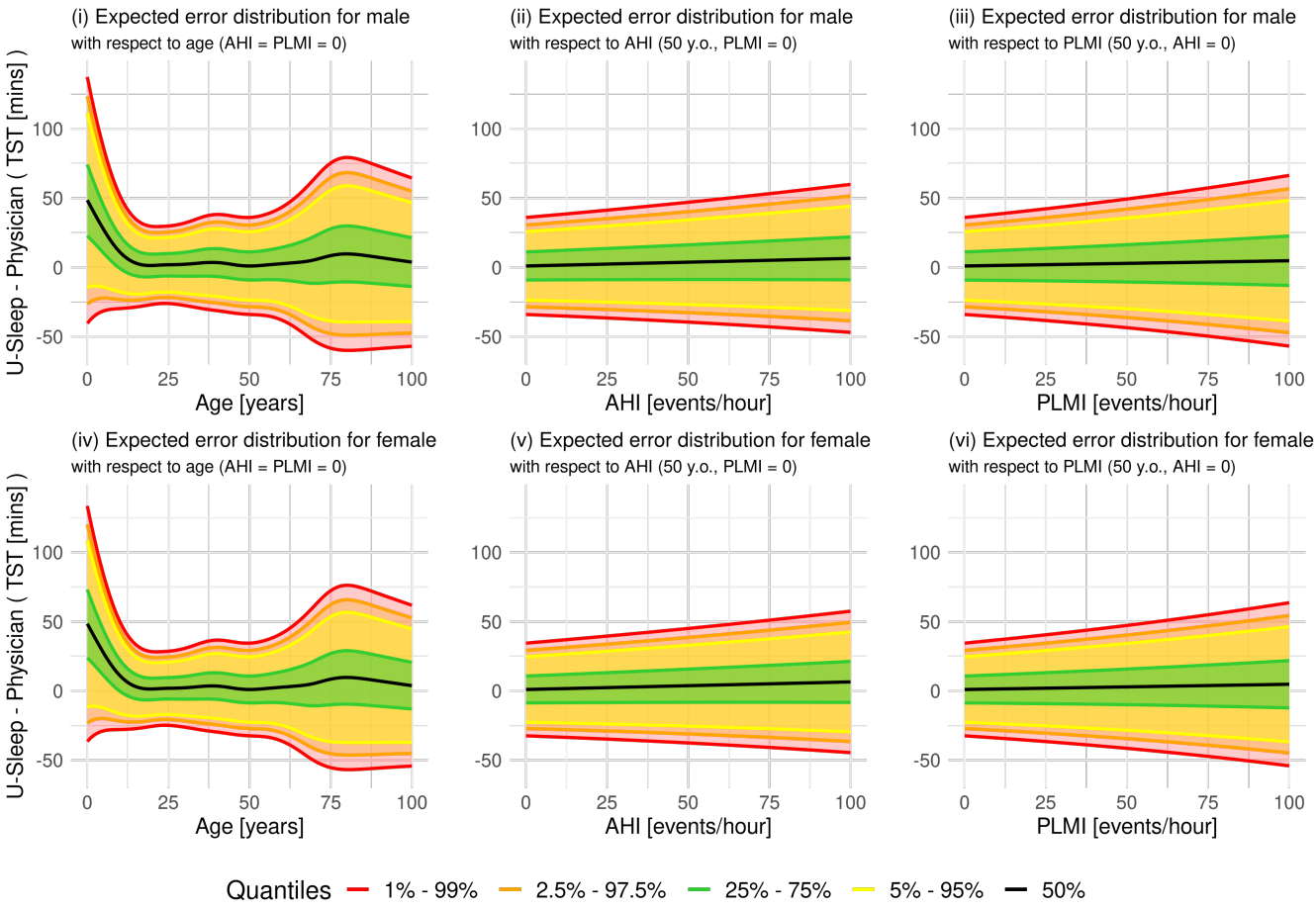

Figure 25. Expected distribution of the bias in the total sleep time (TST, minutes) based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

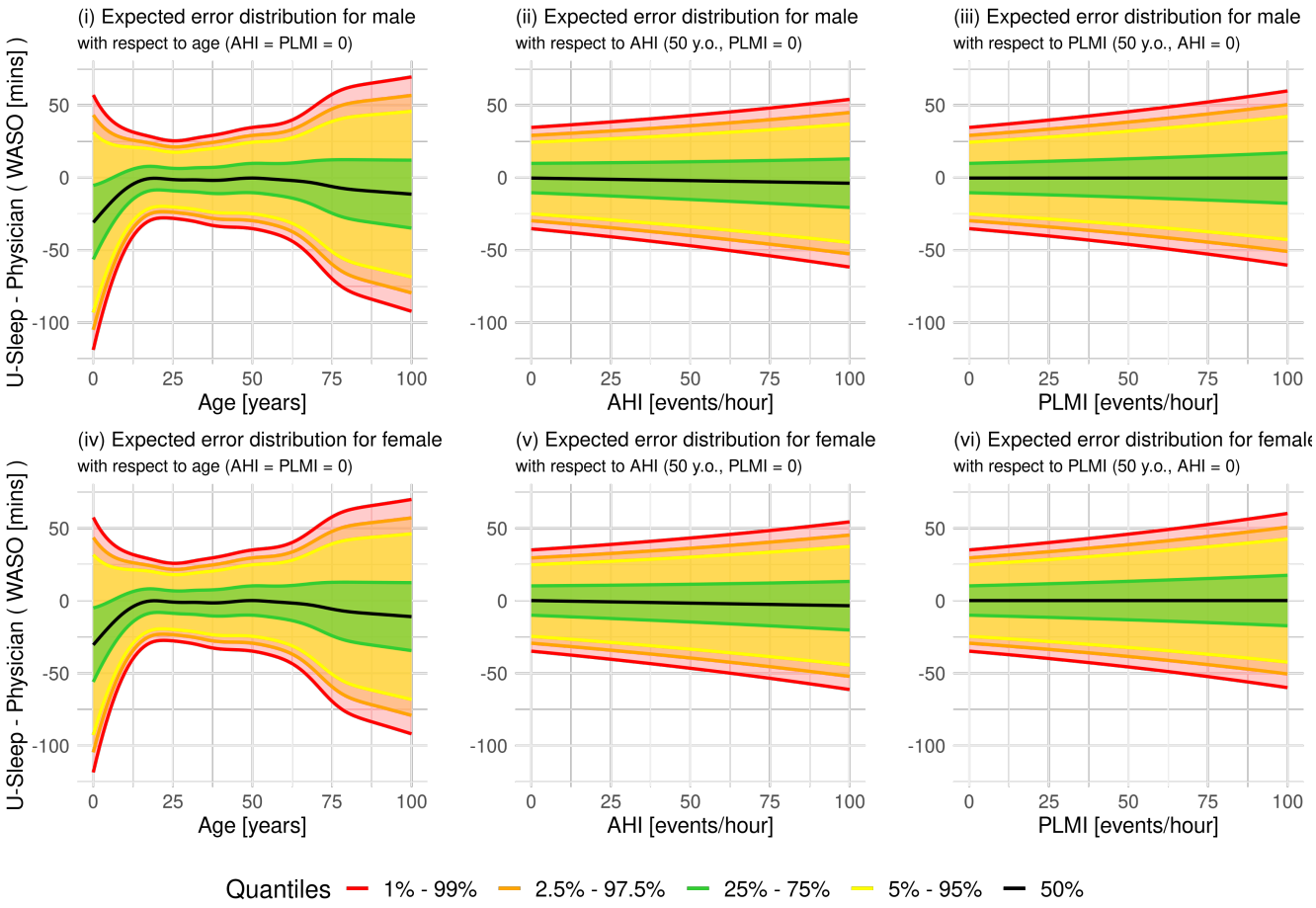

Figure 26. Expected distribution of the bias in the wake after sleep onset (WASO, minutes) based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

Figure 27. Expected distribution of the bias in the number (#) of sleep cycles based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

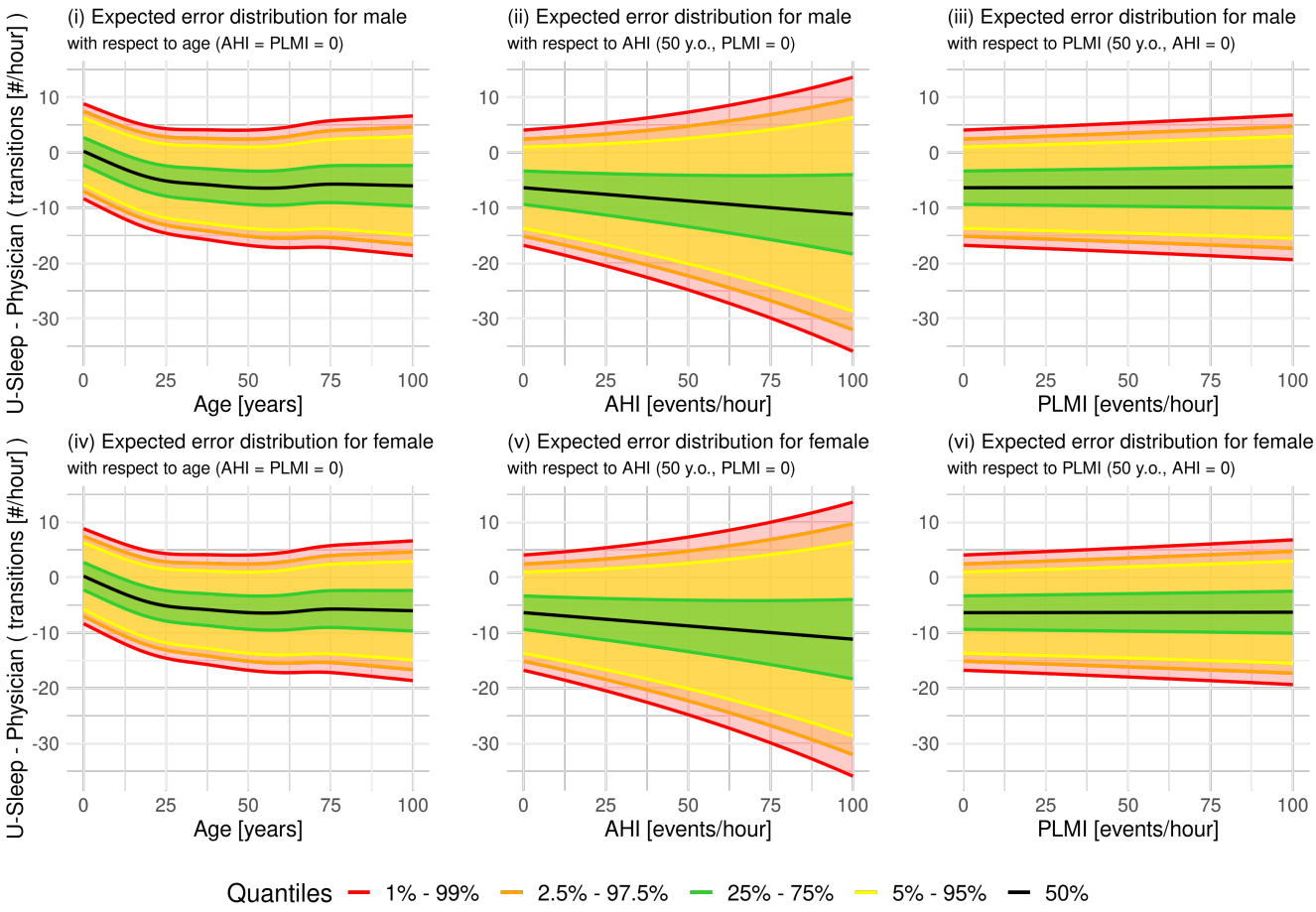

Figure 28. Expected distribution of the bias in the hourly rate (# / hour) of sleep stage transitions based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

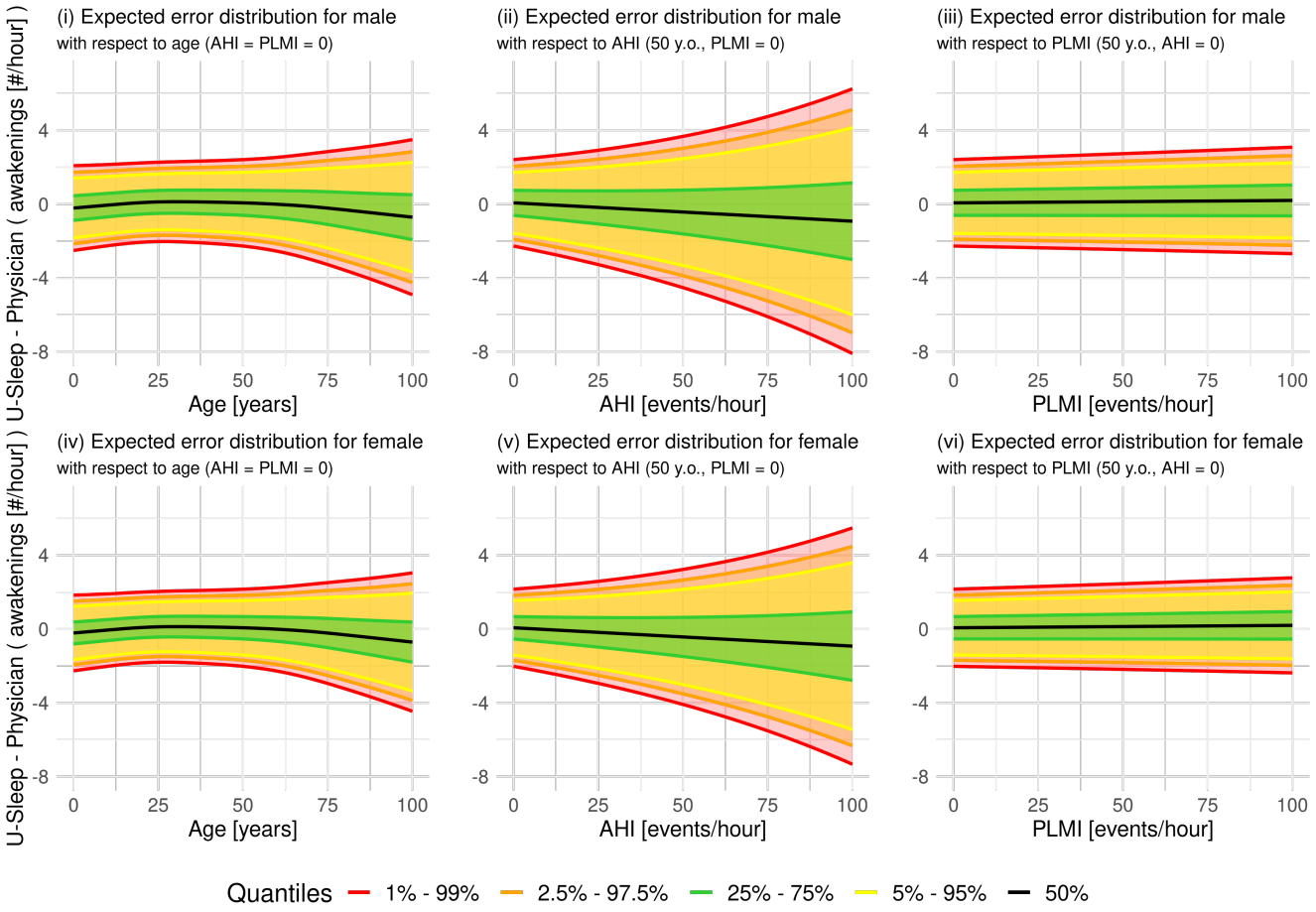

Figure 29. Expected distribution of the bias in the hourly rate (# / hour) of awakenings based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

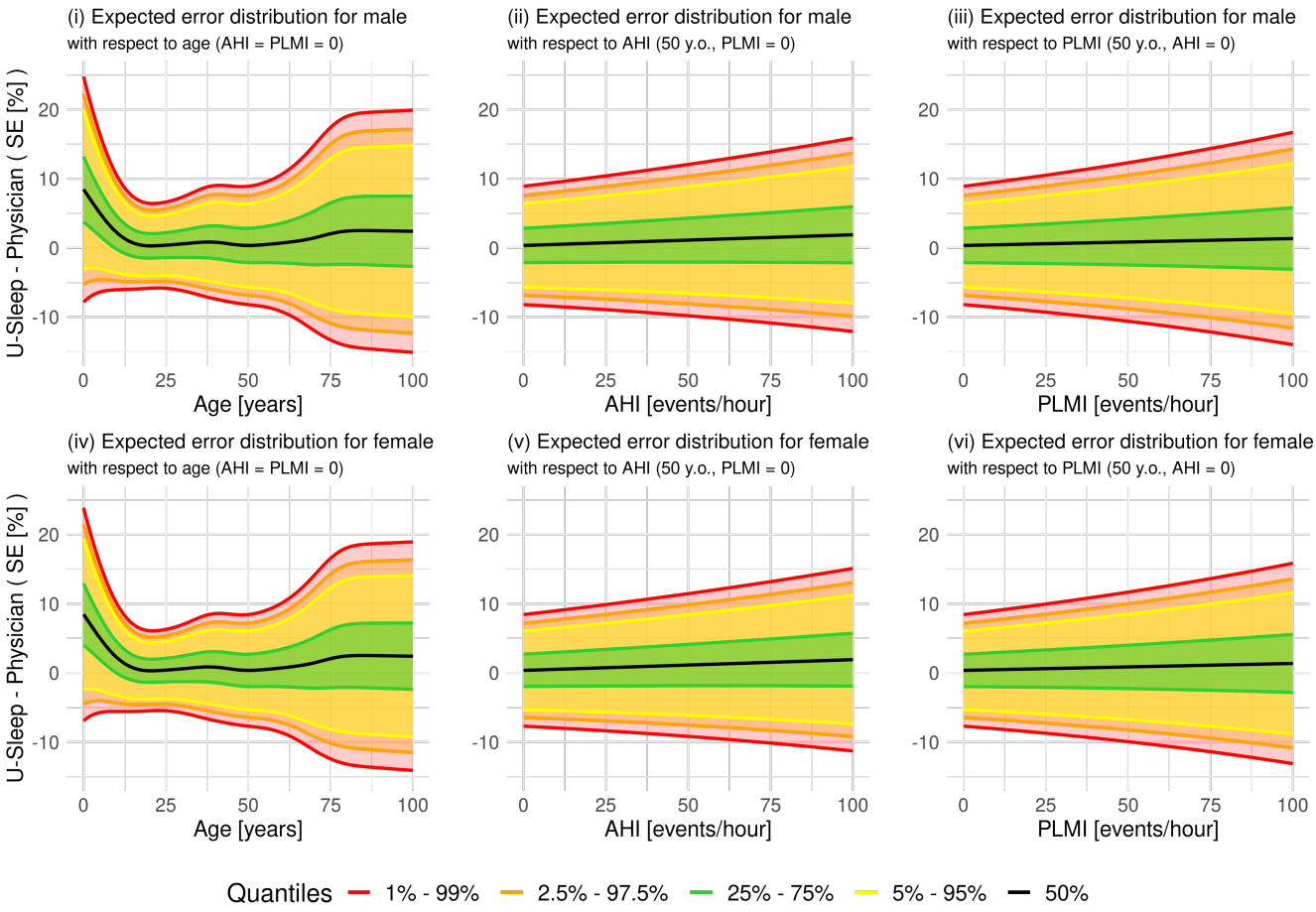

Figure 30. Expected distribution of the bias in the sleep efficiency percentage (SE, %) based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

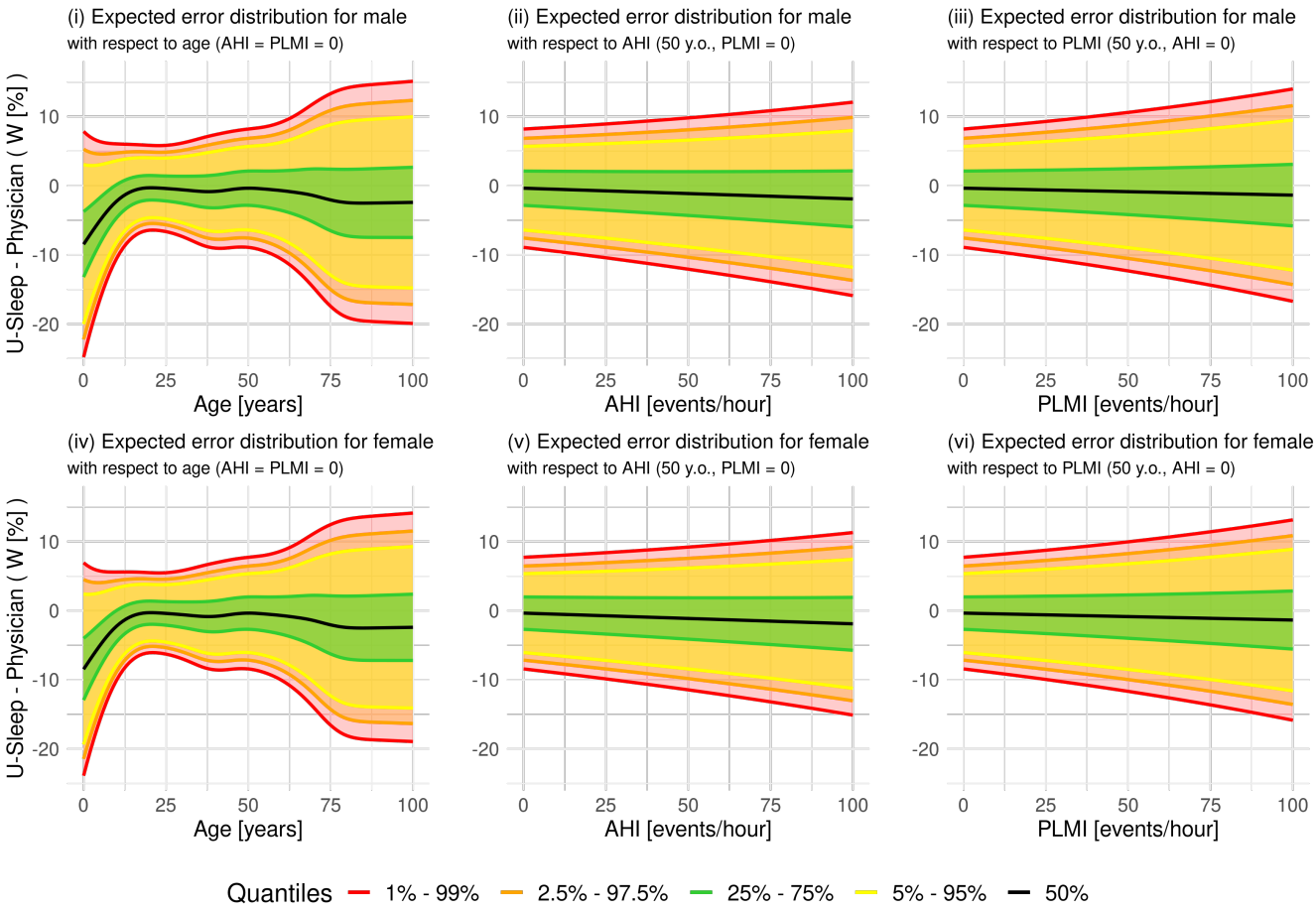

Figure 31. Expected distribution of the bias in the wakefulness percentage after sleep onset (W, %) based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

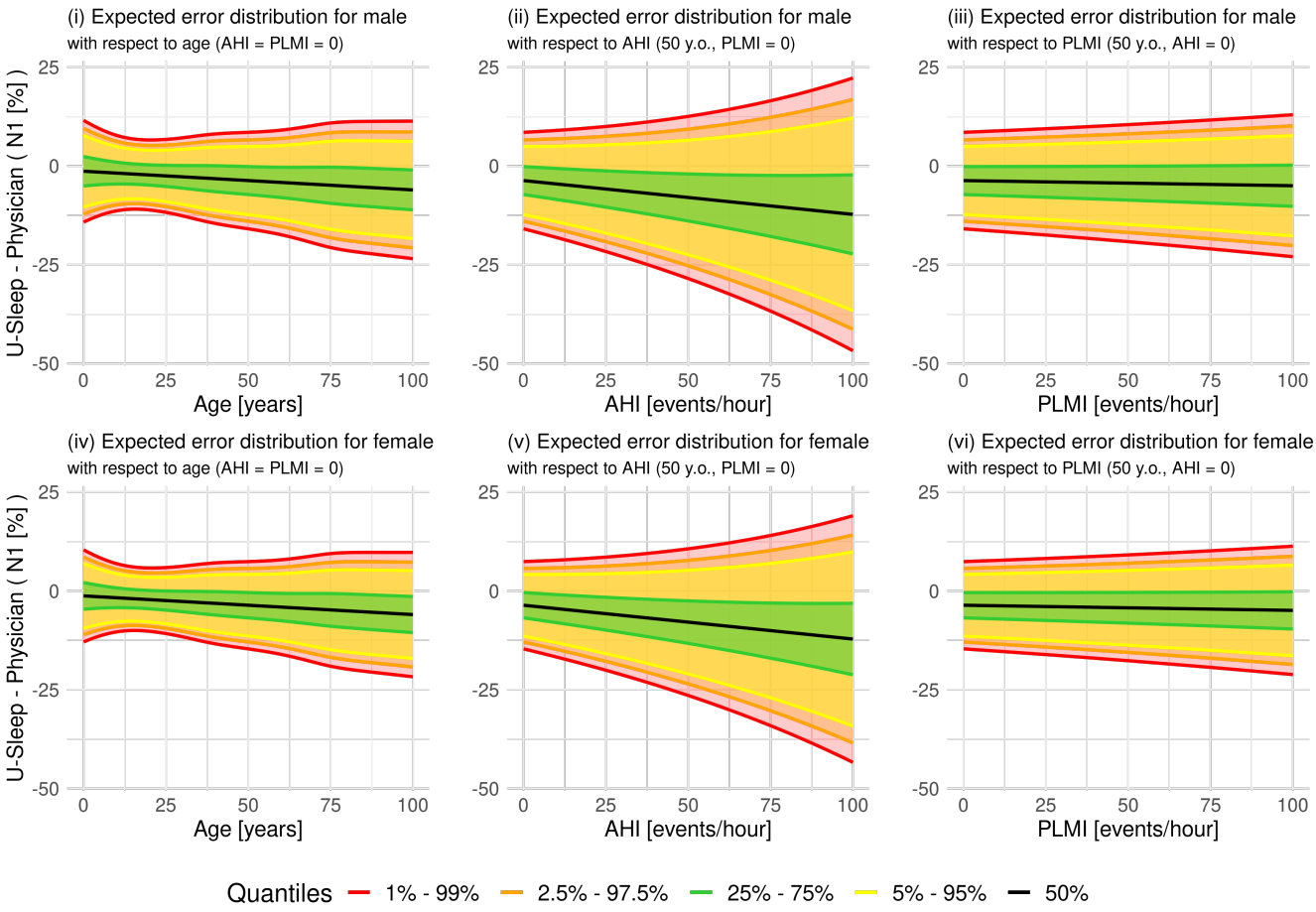

Figure 32. Expected distribution of the bias in the N1 sleep percentage after sleep onset (N1, %) based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

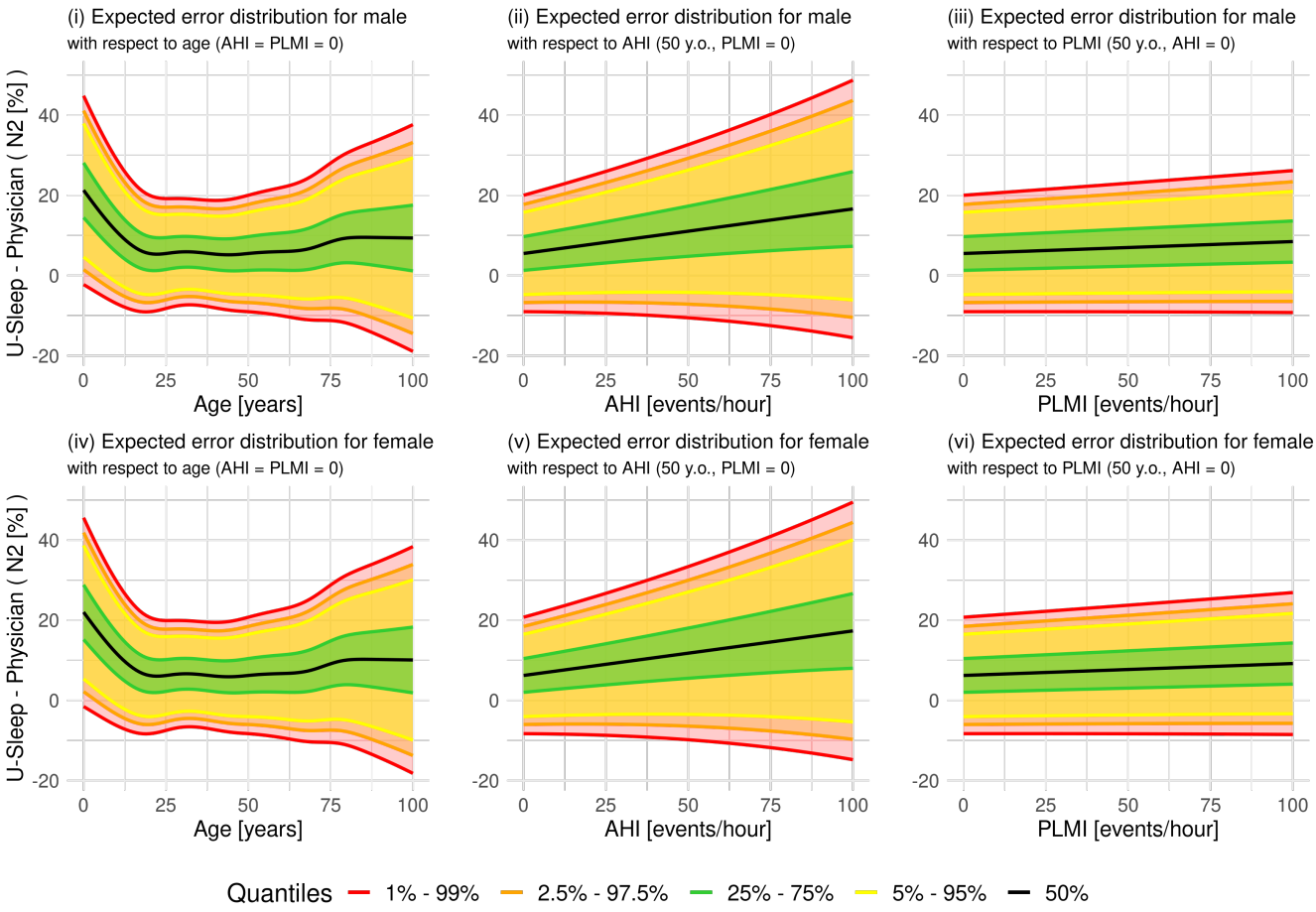

Figure 33. Expected distribution of the bias in the N2 sleep percentage after sleep onset (N2, %) based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

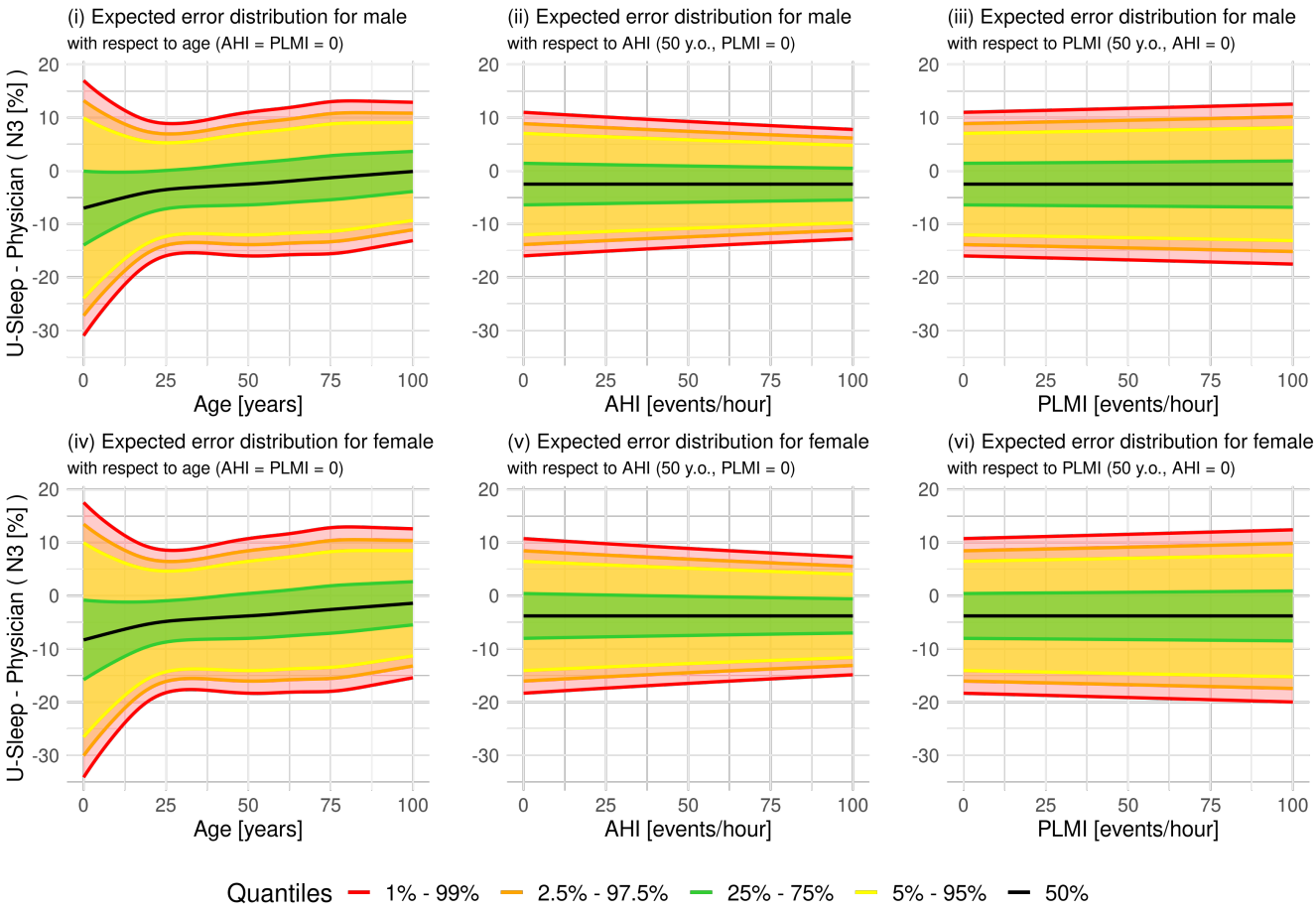

Figure 34. Expected distribution of the bias in the N3 sleep percentage after sleep onset (N3, %) based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.

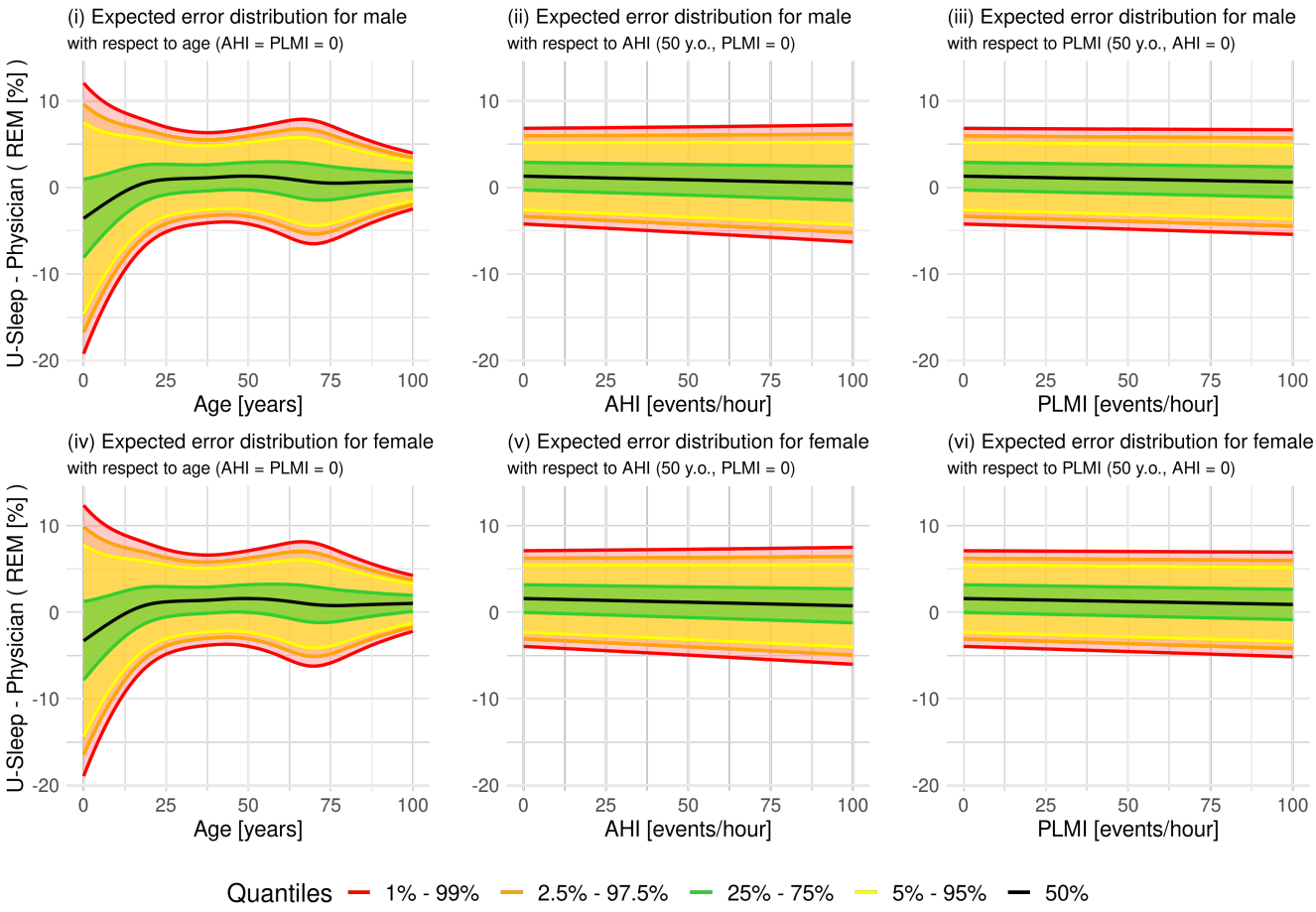

Figure 35. Expected distribution of the bias in the REM sleep percentage after sleep onset (REM, %) based on the generalized normal distribution for U-Sleep predictions, stratified by gender (top row: males, bottom row: females). The graphs display the estimated distribution as a function of bias-inducing variables (age, AHI, and PLMI) on the horizontal axis. The solid black line represents the median, while the shaded areas correspond to different percentile ranges: 25-75% in green, 5-95% in yellow, 2.5-97.5% in orange, and 1-99% in red, illustrating the expected performance variability across subjects' characteristics.