

Módulo 2. Evaluación del rendimiento de la tecnología

Utilizar la tecnología del sueño como ejemplo

Algoritmos

Los datos de actigrafía generalmente se califican y registran en intervalos de un minuto, proporcionando un valor de actividad para cada minuto. Los primeros actogramas publicados resumieron los datos por hora utilizando la actividad de dos pacientes psiquiátricos hospitalizados: uno con anorexia nerviosa y otro con depresión psicótica. Los investigadores pudieron discernir rápidamente patrones de sueño y vigilia, y aunque la precisión fue alrededor del 80% de la medida obtenida a través de PSG (polisomnografía), proporcionó un buen punto de partida para los dispositivos actigráficos portátiles utilizados para medir los períodos de sueño.

La primera técnica de puntuación informó sobre el uso de recuentos de movimiento en bruto del EEG y la actigrafía (Kupfer et al., 1972). Los evaluadores que utilizaban el sistema MediLog usaban "la mejor suposición que se podía derivar" (Kripke et al., 1978, p. 85). Para 1980, se consideraba que la puntuación era el resultado de la mejor suposición utilizando dos evaluadores para minimizar el error (Mullaney et al., 1980).

La necesidad de desarrollar un algoritmo de puntuación que pudiera operacionalizar y automatizar este proceso se hizo evidente. Si bien la actigrafía no podía reemplazar la polisomnografía, podía proporcionar una alternativa más económica, fácil y ligera para capturar datos objetivos de sueño.

Un algoritmo efectivo debe:

- Incorporar datos de varios intervalos de tiempo antes y después del que se está evaluando para determinar la probabilidad de sueño frente a vigilia.
- Tener en cuenta la hora del día para fortalecer la predictibilidad del sueño frente a la vigilia.
- Distinguir entre el sueño y la vigilia tranquila en ausencia de movimiento.
- Tener en cuenta el tiempo en que el dispositivo no está en la muñeca, como durante el baño o el ejercicio.
- Situar cambios breves en los niveles de actividad dentro de un contexto más amplio.
- Tener en cuenta las limitaciones propias del dispositivo.

Los primeros algoritmos solo consideraban niveles generales de actividad, asumiendo que una actividad alta ocurría durante la vigilia y una actividad baja ocurría durante el sueño. Sin embargo, también comenzaron a explorar cambios de momento a momento



en la actividad. Se consideraba probable que los períodos de movimiento intercalados con unos minutos de quietud ocurrieran durante la vigilia. Por lo tanto, no toda la quietud es indicativa del sueño. Los algoritmos efectivos utilizan el contexto. Si una persona estaba en movimiento uno o dos minutos antes, esto es más importante para predecir la probabilidad de sueño o vigilia que el movimiento de hace 10 minutos. El contexto de la hora del día también es importante. La ausencia de movimiento durante el día, cuando una persona está normalmente despierta, es poco probable que sea un episodio de sueño.

La primera puntuación de actigrafía se publicó en 1972 y los investigadores la llamaron 'telemetría' (Kupfer et al., 1972). Ocho pacientes tuvieron actividad motora registrada durante 24 horas y se comparó con el EEG. Aunque hubo algunas discrepancias (como se ve en la tabla 1), los datos estaban bien correlacionados.

La palabra 'actigrafía' apareció por primera vez en la literatura del sueño en 1978 (Kripke et al., 1978), aunque ya se había utilizado en el ejército de los Estados Unidos antes de eso. Los investigadores utilizaron intervalos de un minuto para estimar el tiempo en la cama, la duración del sueño y el tiempo de vigilia después de quedarse dormido, lo que demostró una buena concordancia con la PSG (tabla 1). Por lo tanto, en este momento, comenzará a ver las métricas que definen la concordancia entre dispositivos. Llegaremos a eso. Pero, al principio, esencialmente solo tenían coeficientes de correlación. Con el tiempo, se consideró que eso era demasiado laxo, no lo suficientemente preciso, y es ahí donde tenemos otras métricas que usamos hoy en día.

Tabla 1 MediLog versus. PSG

Comparison of EEG and actigraphic scoring ^a		
	EEG scoring (min)	
	Sleep	Wake
Actigraphic scoring (min)		
Sleep	378	24
Wake	8	172

^aAverage for 102 recordings.

Fuente: Mullaney *et al.*, 1980, p. 88.

Comparison of EEG and actigraphic scoring	Comparación entre la puntuación del EEG y la actigrafía
EEG scoring (min)	Puntuación del EEG (min)
Sleep	Sueño
Wake	Vigilia
Actigraphic scoring (min)	Puntuación actigráfica (min)
Average for 102 recordings	Promedio de 102 registros



En 1980, un estudio a gran escala comparó a 39 pacientes con 63 sujetos de control sanos, que resultaron ser estudiantes universitarios y personal hospitalario (Mullaney et al., 1980). El dispositivo MediLog registró intervalos de un minuto y se comparó con PSG, que también se puntuó en intervalos de un minuto (tabla 1).

Las discrepancias ocurrieron cuando el dispositivo registraba el sueño, mientras que el EEG detectaba la vigilia, y en general, en un problema de clasificación errónea que persiste hoy en día, los dispositivos ponibles subestiman la vigilia.

Los datos mostrados en la tabla 2 de este mismo estudio parecían capturar muchos de los obstáculos más desafiantes que surgieron en el campo en las siguientes décadas. El primero de estos fue que los dispositivos ponibles eran más efectivos para capturar el sueño en personas no pacientes. Esto se debe a que el sueño de los pacientes tiende a ser más variable y menos predecible en comparación con las personas no pacientes. Por lo tanto, es más difícil puntuar debido a que suele haber más ruido. Pueden ocurrir escenarios inesperados en un entorno de pacientes.

Tabla 2 Actigrafía vs. EEG

Correlation between actigraph and EEG scoring									
Group variable	TSP		TST		WASO		MSA		Agreement (%)
All subjects (n = 102)	0.90 ^a		0.89 ^a		0.70 ^a		0.25 ^c		94.5
Patients (n = 39)	0.82 ^a		0.81 ^a						
Nonpatients (n = 63)	0.97 ^a	(p < 0.0001)	0.95 ^a	(p < 0.001)	0.56 ^a	(p < 0.001)	0.09 ^e	(p < 0.05)	91.6 (p < 0.001)
					0.87 ^a		0.46 ^a		96.3
Age									
≥50 (n = 17)	0.52 ^a		0.39 ^e		0.82 ^a		0.01 ^e		88.5
<50 (n = 85)	0.95 ^a	(p < 0.0001)	0.95 ^a	(p < 0.0001)	0.65 ^a	(ns)	0.32 ^b	(ns)	95.7 (p < 0.001)
EEG TSP									
≥390 (n = 66)	0.85 ^a		0.83 ^a		0.72 ^a		0.19 ^e		94.7
<390 (n = 36)	0.74 ^a	(ns)	0.78 ^a	(ns)	0.66 ^a	(ns)	0.52 ^a	(ns)	94.2 (ns)
Sleep log TST									
≥390 (n = 53)	0.96 ^a		0.94 ^a		0.72 ^a		0.26 ^d		96.3
<390 (n = 49)	0.79 ^a	(p < 0.0001)	0.69 ^a	(p < 0.0001)	0.70 ^a	(ns)	0.23 ^e	(ns)	92.5 (p < 0.001)
Recording quality									
Low (n = 20)	0.78 ^a		0.79 ^a		0.74 ^a		-0.07 ^e		92.8
High (n = 24)	0.97 ^a	(p < 0.001)	0.97 ^a	(p < 0.001)	0.75 ^a	(ns)	0.59 ^c	(p < 0.05)	95.9 (ns)
Interpretability									
Low (n = 27)	0.77 ^a		0.83 ^a		0.62 ^a		-0.16 ^e		90.4
High (n = 24)	0.97 ^a	(p < 0.001)	0.97 ^a	(p < 0.01)	0.75 ^a	(ns)	0.59 ^c	(p < 0.01)	95.9 (p < 0.005)

^a p < 0.0001 (one-tailed), ^b p < 0.001, ^c p < 0.01, ^d p < 0.05, ^e not significant.
Abbreviations: TSP, total sleep period; TST, total sleep time; WASO, wake after sleep onset; and MSA, number of midsleep awakenings.

Fuente: Mullaney et al., 1980, p. 87.

Correlation between actigraph and EEG scoring	Correlación entre la puntuación de la actigrafía y el EEG
Group variable	Variable del grupo
TSP	TSP
TST	TST
WASO	WASO
MSA	MSA
Agreement	Concordancia
All subjects	Todos los sujetos
Patients	Pacientes
Nonpatients	Personas no pacientes



Age	Edad
EEG TSP	EEG
Sleep log TST	Registro de sueño TST
Recording quality Low High	Calidad de grabación Baja Alta
Interpretability	Interpretabilidad
Not significant	No significativa
Abbreviations: TSP, total sleep period; TST, total sleep time; WASO, wake after sleep onset; and MSA, number of midsleep awakenings	Abreviaturas: TSP, período de sueño total; TST, tiempo total de sueño; WASO, tiempo de vigilia después del inicio del sueño; y MSA, número de despertares a mitad del sueño

En segundo lugar, los dispositivos eran más capaces de medir el movimiento durante el sueño en personas más jóvenes, lo que daba lugar a un sesgo relacionado con la edad.

En tercer lugar, había un sesgo con la duración total del sueño. Aquellos con un sueño más corto tenían menos intervalos de tiempo totales. Estos resultan en un mayor peso en cualquier intervalo en particular. Esto significa que puntuar incorrectamente un intervalo en una persona que duerme un tiempo corto puede sesgar los resultados en mayor medida que en alguien que duerme más tiempo. Por lo tanto, aquellos que dormían más de seis horas y media (390 minutos) tenían una mayor precisión.

Por último, la calidad de grabación y la interpretabilidad siguen siendo problemas hasta el día de hoy. Los registros difíciles tienen más probabilidades de resultar en desacuerdo entre los evaluadores. Lo mismo sucede con un registro de EEG de baja calidad.

Después de descubrir que los registros actigráficos estaban coincidiendo bastante con PSG en una base minuto a minuto, los investigadores comenzaron a cuestionarse si la puntuación podría automatizarse. Hicieron esto ingresando los siguientes factores en una ecuación de predicción:

1. Actividad total en el intervalo actual
2. La muestra más activa del minuto
3. La suma de las dos muestras más activas separadas por más de 30 segundos
4. La suma de las ocho muestras más activas
5. Pesos para la actividad total en los cuatro intervalos anteriores y los dos intervalos posteriores para tener en cuenta el contexto.

Después de experimentar con este paradigma, demostraron que se podían eliminar (1), (3) y (4). Por lo tanto, la muestra más activa del minuto y la actividad total en los cuatro intervalos anteriores y los dos siguientes fueron los que tuvieron la mayor precisión en predecir el comportamiento sueño-vigilia en cualquier intervalo en particular. Kripke y colegas (1978) utilizaron 20 registros, desarrollaron un algoritmo basado en 17 de ellos y lo probaron en tres individuos.



El algoritmo de Webster, con 'I' siendo el intervalo de interés, con cuatro intervalos anteriores y dos posteriores.

Algoritmo de Webster

$$D = 0.025 \times [0.15T_{(i-4)} + 0.15T_{(i-3)} + 0.15T_{(i-2)} + 0.08T_{(i-1)} + 0.21T_{(i)} + 0.12T_{(i+1)} + 0.13T_{(i+2)}]$$

Esto resultó en un acuerdo del 94,46% en el grupo de desarrollo y del 96,02% en las tres personas que conformaron la muestra de validación. Sobrevaloró el sueño más de lo que sobrevaloró la vigilia, como ocurrió en estudios anteriores. Esto se debe a que el inicio del sueño ocurre después de que cesa el movimiento. Aquí es donde el uso del movimiento por sí solo tiene sus limitaciones. Sin embargo, el algoritmo ofreció una mejor predicción del sueño frente a la vigilia al incorporar la actividad de los intervalos anteriores al modelo. La falta de movimiento precedida por una falta de movimiento tenía más probabilidades de ser un registro de sueño, por ejemplo.

Se crearon algunas reglas de seguimiento para abordar el problema de informar en exceso el sueño:

1. si tienes de 4 a 9 minutos de vigilia, recodifica el primer minuto de sueño como vigilia.
2. Si tienes de 10 a 14 minutos de vigilia, recodifica los primeros 3 minutos de sueño como vigilia.
3. si tienes +15 minutos de vigilia, recodifica los primeros 4 minutos de sueño como vigilia.
4. Si tienes de 10 a 19 minutos de vigilia alrededor de un período de sueño, cualquier período de sueño de 6 minutos o menos debe recodificarse como vigilia.
5. Si tienes de 20 minutos de vigilia alrededor de un período de sueño, cualquier período de sueño de 10 minutos o menos debe recodificarse como vigilia.

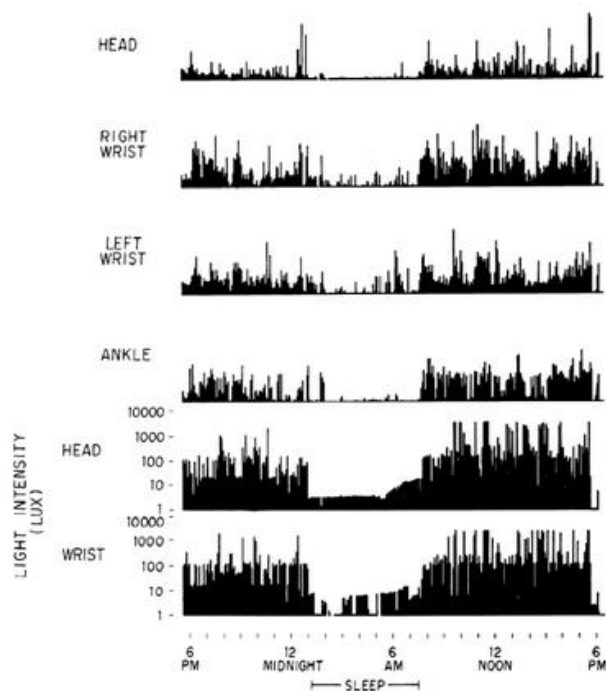
Luego, el algoritmo de Webster se modificó ligeramente para su uso en un dispositivo de grabación portátil. El VitaLog seguía siendo de muñeca, pero el grabador externo se llevaba en un cinturón. Este era un transductor diferente, con electrónica y un amplificador diferente. A pesar de la tecnología mejorada, el algoritmo de Webster todavía resultó en un 93% de precisión con PSG.

La siguiente iteración de este trabajo ocurrió en 1983 cuando se utilizaron grabadores VitaLog en la frente, la muñeca no dominante y el tobillo (Okudaira et al., 1983). Los investigadores también registraron los niveles de luz ambiental a través de un transductor de luz portátil separado, que sirvió como marcador circadiano de movimiento. Muchos dispositivos ponibles para consumidores no registran la luz; sin embargo, los dispositivos de investigación utilizados en un entorno de laboratorio aún registran este parámetro importante.



Como se ve en la imagen 1, la ubicación más efectiva para los dispositivos ponibles de registro de luz fue el tobillo, ya que las ubicaciones como la muñeca son menos propensas a estar cubiertas por la ropa de cama durante la noche. Aunque la cabeza es aún más propensa a estar expuesta a la luz durante la noche, no tiende a moverse tanto como las extremidades, por lo que los resultados tienden a ser más confiables que los dispositivos usados en la muñeca.

Imagen 1. Ubicación de los actígrafos



Fuente: Okudaira *et al.*, 1983.

Right wrist	Cabeza
Left wrist	Muñeca derecha
Ankle	Muñeca izquierda
Head	Tobillo
Wrist	Cabeza
Light intensity	Muñeca
Midnight	Intensidad de la luz
Sleep	Medianoche
Noon	Sueño
Pm	Mediodía
Right wrist	Pm

El siguiente avance ocurrió en 1989 con el algoritmo de Sadeh, cuando el especialista en sueño pediátrico Avi Sadeh investigó el potencial clínico de la actigrafía (Sadeh et al., 1991). Este algoritmo se desarrolló utilizando un proceso similar y encontró una buena concordancia minuto a minuto.

- Controles (edad 20-76; n = 13): 90,2%.
- Apnea del sueño (edad 20-76; n = 24): 85,7%.



- Insomnio (edad 20-76; n = 16): 78,2%.
- Niños (edad 3-13; n = 13): 89,9%.

Esta muestra diversa tuvo menos acuerdo con PSG, pero fue más flexible para pacientes con trastornos del sueño y todavía es el estándar de oro entre las poblaciones pediátricas en la actualidad.

El algoritmo de Sadeh tuvo un rendimiento ligeramente mejor para el insomnio que el de Webster, pero estadísticamente, los dos no diferían significativamente entre sí ni de la puntuación manual. La puntuación manual difería de PSG en aproximadamente 50 minutos, con el 25% de los registros con discrepancias de más de una hora en comparación con PSG. El algoritmo de Webster difería de PSG en aproximadamente 63 minutos, con discrepancias de más de una hora el 45% del tiempo. El algoritmo de Sadeh también difería de PSG en aproximadamente una hora, con un tercio de las muestras con discrepancias de más de 60 minutos.

Luego se compararon los dispositivos actigráficos con PSG para evaluar su capacidad para detectar el sueño frente a la vigilia en pacientes con insomnio primario y secundario. El insomnio primario, o psicofisiológico, ocurre cuando los síntomas de insomnio surgen sin una causa conocida. Por otro lado, el insomnio secundario proviene de una enfermedad médica primaria, trastorno mental u otro trastorno del sueño (Morgenthaler et al., 2006).

Aunque el error promedio, en comparación con PSG para detectar el sueño frente a la vigilia, fue de 49 minutos, se descubrió que los datos de actigrafía se alineaban mejor con las estimaciones del diario de sueño en cuanto a la duración del sueño que PSG, lo que podría ser más adecuado para capturar el insomnio. Hay muchas razones por las que PSG puede ser problemático para diagnosticar el insomnio. Principalmente, los pacientes a menudo están despiertos mientras un PSG informa de que están dormidos. En los diagnósticos de insomnio, los diarios de sueño tienden a ser la medida principal.

En comparación con PSG, un diario de sueño suele reportar muchos menos despertares por noche. Esto se debe a que, aunque los microdespertares ocurren a menudo durante toda la noche, la mayoría de las personas no los recuerdan. Un diario de sueño típico registrará de uno a tres despertares, mientras que un PSG podría informar 10, 20, o incluso más. La actigrafía tiende a subestimar los despertares nocturnos y, por lo tanto, en casos de insomnio, la actigrafía puede ser más confiable que PSG, además de evitar el problema común de la aversión a intentar dormir en un entorno de laboratorio.

El siguiente desarrollo ocurrió cuando el algoritmo de Webster se refinó en 1992 (Cole et al., 1992). El algoritmo Cole-Kripke sigue siendo el estándar de la actigrafía hasta la fecha. El estudio incluyó a 41 sujetos, 15 de los cuales eran controles, junto con varios pacientes psiquiátricos, de apnea del sueño, dolor e insomnio en un intento por diversificar la muestra, al igual que había hecho Sadeh.



Este fue el primer estudio de validación realizado en un dispositivo disponible comercialmente (Motionlogger), cuyas versiones más nuevas todavía están disponibles para su compra en la actualidad. El dispositivo original muestreaba 30 veces por segundo utilizando recuentos de movimiento de dos segundos. Evaluaba la actividad promedio y los periodos máximos de 2, 6, 10, 20 y 30 segundos dentro de cada intervalo. Descubrieron que el período máximo de 10 segundos era el que tenía una mayor capacidad predictiva. El algoritmo Cole-Kripke, con 'I' siendo el intervalo de interés, utiliza los cuatro intervalos anteriores al intervalo en cuestión y los dos que le siguen para puntuar el sueño frente a la vigilia. Si $D > 1$, el intervalo se puntúa como vigilia.

Algoritmo de Cole-Kripke

$$D = 0.00001 \times [404T_{(i-4)} + 598T_{(i-3)} + 326T_{(i-2)} + 441T_{(i-1)} + 1408T_{(i)} + 508T_{(i+1)} + 350T_{(i+2)}]$$

Es importante mencionar que este algoritmo se desarrolló utilizando un transductor de movimiento, y a pesar del salto a la actigrafía en dispositivos modernos, este algoritmo todavía se aplica. Se utiliza el intervalo de interés junto con los cuatro intervalos anteriores y los dos que le siguen para predecir el sueño frente a la vigilia.

Este algoritmo resultó en una concordancia del 87,05% en el grupo de desarrollo y del 87,91% en la muestra de validación, y aumentó a 87,93% y 88,25% cuando se aplicaron las reglas de seguimiento de Webster. Los parámetros de sueño durante toda la noche se pueden ver en la tabla 2.

Tabla 3 Cole-Kripke

	Training sample			Validation sample		
	PSG ^b	Actigraph ^b	<i>r</i> ^c	PSG	Actigraph	<i>r</i>
Minutes scored	427.0 ± 74.7	427.0 ± 74.7		445.5 ± 43.1	445.5 ± 43.1	
Total sleep time (minutes) ^d	308.0 ± 93.5	329.5 ± 95.1*	0.91***	344.5 ± 52.5	363.9 ± 54.4*	0.77***
Percent sleep (%) ^{d,e}	71.4 ± 20.3	76.7 ± 20.1*	0.89***	77.9 ± 13.0	82.4 ± 13.5*	0.82***
Sleep efficiency (%) ^{d,e}	80.4 ± 18.7	84.4 ± 19.4	0.85**	85.0 ± 11.6	88.6 ± 11.4	0.71**
Sleep latency (minutes) ^f	75.6 ± 89.7	70.9 ± 102.2	0.94***	59.2 ± 46.1	50.1 ± 50.7	0.90***
Wake time after sleep onset (minutes) ^f	51.3 ± 45.6	40.2 ± 32.7	0.49*	49.9 ± 37.2	36.6 ± 30.6*	0.63**

Fuente: Cole *et al.*, 1992, p. 465.

Training sample	Muestra de entrenamiento
Validation sample	Muestra de validación
Actigraph	Actigrafía
PSG	PSG
Minutes scored	Minutos puntuados
Total sleep time (minutes)	Tiempo de sueño total (minutos)
Percent sleep	Porcentaje de sueño
Sleep efficiency	Eficiencia del sueño
Sleep latency (minutes)	Latencia del sueño (minutos)
Wake time after sleep onset (minutes)	Tiempo de vigilia luego del inicio del sueño (minutos)



^a La puntuación se realizó utilizando el algoritmo para un intervalo máximo de 10 segundos que se superpone por minuto. La re-puntuación se basó en las reglas de seguimiento de Webster.

^b Medias + DE. Los asteriscos indican diferencias significativas entre PSG y actigrafía mediante la prueba *t* pareada.

^c Correlación de Pearson entre las puntuaciones de PSG y actigrafía. Los asteriscos indican correlaciones significativas.

^d Incluye todos los minutos puntuados como sueño, incluso aquellos que ocurrieron antes de que se cumpliera el criterio de 20 minutos para el inicio del sueño.

^e El porcentaje de sueño se basa en todo el registro. La eficiencia del sueño se basa en el tiempo en la cama.

^f Tanto el inicio del sueño de PSG como de actigrafía se definen como el comienzo del primer intervalo que contiene 20 minutos puntuados como sueño, sin más de un minuto de vigilia de por medio. Las latencias se calcularon desde el comienzo de la grabación simultánea de actigrafía/PSG, independientemente de si el sujeto intentaba dormir. * $p < 0,05$. ** $p < 0,002$. *** $p < 0,001$

Los diferentes transductores funcionan de diferentes maneras, pero los algoritmos antiguos se han aplicado a datos más nuevos con un éxito sorprendente. Cuando el algoritmo de Webster, que se calculó utilizando el transductor de movimiento original, se aplicó a los datos del Motionlogger de Cole-Kripke, hubo concordancias que solo se desplazaron ligeramente de un 86,44% en la muestra de entrenamiento y un 87,54% en la muestra de validación a un 87,24% y 87,73%, respectivamente.

El primer artículo de revisión de la actigrafía, publicado en 1995, resumió los estudios realizados hasta la fecha y consolidó las tasas de concordancia con PSG (Sadeh et al., 1995).

Tabla 4 Resumen de las tasas de concordancia de la actigrafía hasta 1995



Study ^a	Sample ^b	Sample size	Age	S/W ^c	SEF ^c	DUR ^c	Comments ^d
Kripke (3)	N	5	NA	NA	0.98	0.95	
Mullaney (4)	N	53	18-66 yr	96.3	0.81	0.97	HS TST
	P	32	18-66 yr	91.6	0.95	0.82	HS TST
Webster (5)	N	14	College	93.9	NA	NA	AS IDT
	N + P	14	College	93.4	NA	NA	AS IDT
Sadeh (6)	N	13	20-76 yr	90.2	0.91	NA	AS
	SAS	25	20-76 yr	85.7	0.63	NA	AS
	INS	16	20-76 yr	78.2	0.79	NA	AS
	P/C	13	3-13 yr	89.9	0.81	NA	AS
Sadeh (7)	N + P/C	11	12-48 mo	85.3	NA	NA	AS
Hauri (8)	INS	36	24-69 yr	82.1	NA	NA	AS
Cole (11)	N + P	51	NA	88.0	0.82	0.90	AS SLT
Sadeh (12)	N	36	10-25 yr	91.2	NA	NA	AS
Sadeh (13)	N	41	Newborn-12 mo	95.3	NA	0.95	AS

Fuente: Sadeh *et al.*, 2005, p. 290.

Study	Estudio
Sample	Muestra
Sample size	Tamaño de la muestra
Age	Edad
S/W	S/W
SEF	SEF
DUR	DUR
Comments	Comentarios
Kripke	Kripke
Mullaney	Mullaney
Webster	Webster
Sadeh	Sadeh
Hauri	Hauri
Cole	Cole
Sadeh	Sadeh
College	Universidad
NA	NA

En 2001, Jean-Louis y sus colegas desarrollaron el algoritmo de Jean-Louis y utilizaron el dispositivo Actillum, que fue el primero en incluir acelerometría. Si bien los pesos dados a los intervalos que rodean al que se cuestiona eran ligeramente diferentes a los algoritmos que existían previamente, este fue el primer algoritmo diseñado para un acelerómetro en lugar de un transductor de movimiento, y tuvo una concordancia del 87,1% con PSG, que aumentó a 88,3% después de aplicar las reglas de seguimiento de Webster.

El algoritmo de Jean Louis

$$D = 0.13 \times [0.010T_{(i-4)} + 0.015T_{(i-3)} + 0.028T_{(i-2)} + 0.31T_{(i-1)} + 0.085T_{(i)} + 0.015T_{(i+1)} + 0.010T_{(i+2)}]$$

Muchos otros dispositivos aparecieron en el mercado en la década de 2000, con tasas de concordancia relativamente similares, incluidos el Actiwatch, Sleepwatch, Actical, Actiwatch-L y Spectrum.

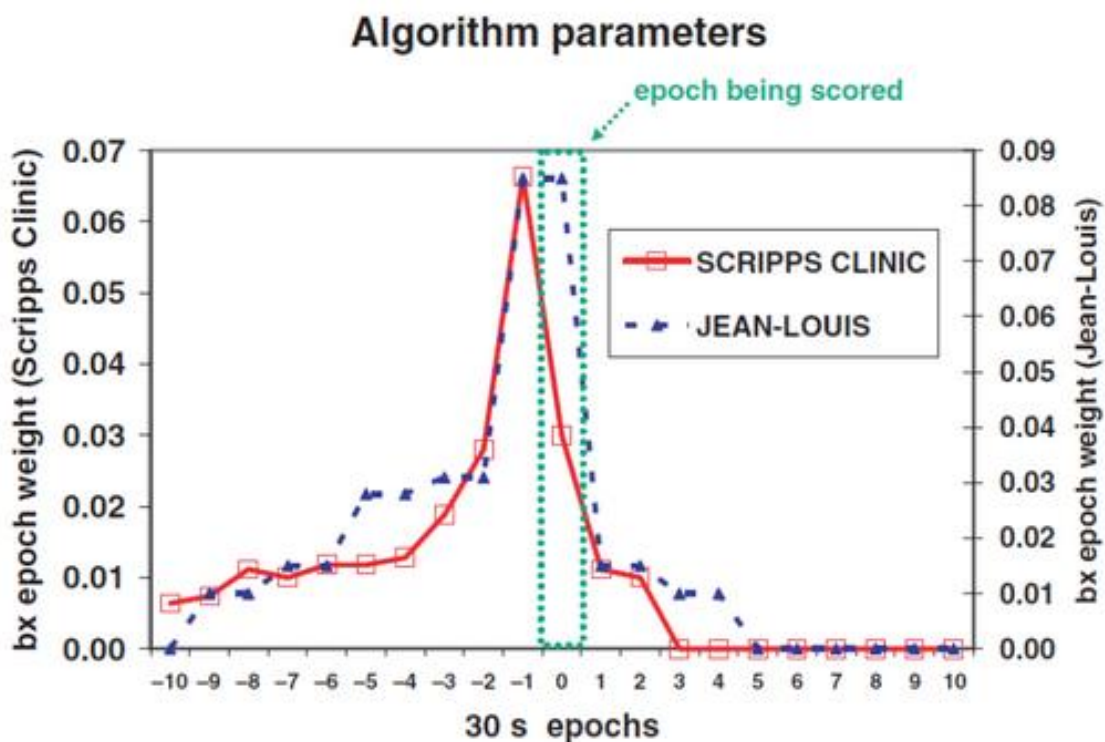


Para el año 2010, el Actiwatch-L era uno de los dispositivos más utilizados, pero el algoritmo no se hizo público. Kripke y sus colegas avanzaron en los métodos anteriores en una muestra de 49 individuos y en intervalos de 30 segundos para alinearse con PSG (Kripke et al., 2010). El nuevo algoritmo Scripps incorporó 10 intervalos anteriores y dos posteriores, lo que resultó en una concordancia del 87% con PSG. La concordancia con PSG para el Actiwatch Spectrum fue ligeramente más baja del 85%, la razón sigue siendo desconocida.

Algoritmo de Scripps

$$D = 0.30 \times [0.0064T_{(i-10)} + 0.0074T_{(i-9)} + 0.0112T_{(i-8)} + 0.0112T_{(i-7)} + 0.0118T_{(i-6)} + 0.0118T_{(i-5)} + 0.0128T_{(i-4)} + 0.0188T_{(i-3)} + 0.0280T_{(i-2)} + 0.0664T_{(i-1)} + 0.0300T_{(i)} + 0.0112T_{(i+1)} + 0.1000T_{(i+2)}]$$

Imagen 2. Algoritmo de Scripps versus algoritmo de Jean-Louis



Fuente: Kripke *et al.*, 2010.

Algorithm parameters	Parámetros del algoritmo
Epoch being scored	Intervalo puntuado
Bx epoch weight (Scripps clinic)	Ponderación del intervalo Bx (Clínica Scripps)
Scripps clinic	Clínica Scripps
Jean-louis	Jean-louis
epochs	intervalos

En comparación con el algoritmo de Jean-Louis, el algoritmo Scripps estaba ligeramente ponderado hacia el intervalo anterior al actual, pero ambos están alineados de manera similar, ambos parecen estabilizarse entre el 85% y el 90% como resultado de limitaciones de rendimiento inherentes (Kripke et al., 2010).

Rendimiento

El rendimiento actigráfico se evalúa a través de tres parámetros vitales: sensibilidad, especificidad y precisión.

- Sensibilidad: el porcentaje de intervalos de "verdadero" sueño puntuadas como "sueño".
- Especificidad: el porcentaje de intervalos de "verdadera" vigiliadas puntuadas como "vigilia".
- Precisión: el porcentaje de intervalos en las que la puntuación del dispositivo coincide con la grabación PSG.

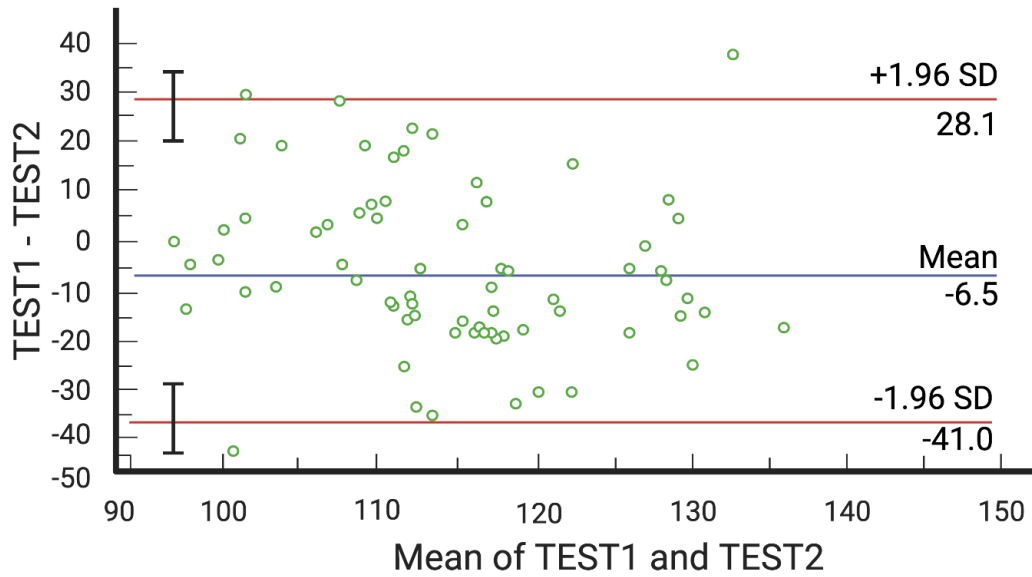
La sensibilidad suele ser alta si la mayoría del episodio de grabación es de sueño. La especificidad es a menudo la métrica más débil, ya que refleja la incapacidad para detectar las épocas de vigilia que ocurren en medio de un período de sueño. Esto ha surgido como la verdadera prueba de cuán efectivo es un dispositivo. Aunque la precisión es una métrica crítica clave, la concordancia puede estar sesgado en ciertas situaciones, como en pacientes con insomnio, y esto debe tenerse en cuenta al seleccionar el mejor dispositivo para un individuo.

También, normalmente, se informan los puntajes kappa en estudios que evalúan el rendimiento de dispositivos ponibles, ya sea para el sueño, la actividad física o la tecnología portátil de rendimiento. Esto representa el acuerdo entre dos "evaluadores" o "puntuadores" de los datos utilizados para dos metodologías diferentes, es decir, PSG y un reloj de pulsera. Los puntajes van desde 0 (azar) hasta 1 (concordancia perfecta) y se utiliza la siguiente ecuación:

$$\kappa = (\text{observado} - \text{esperado}) / (1 - \text{esperado})$$

Imagen 3. Ejemplo de la gráfica Bland-Altman

Bland-Altman Plot



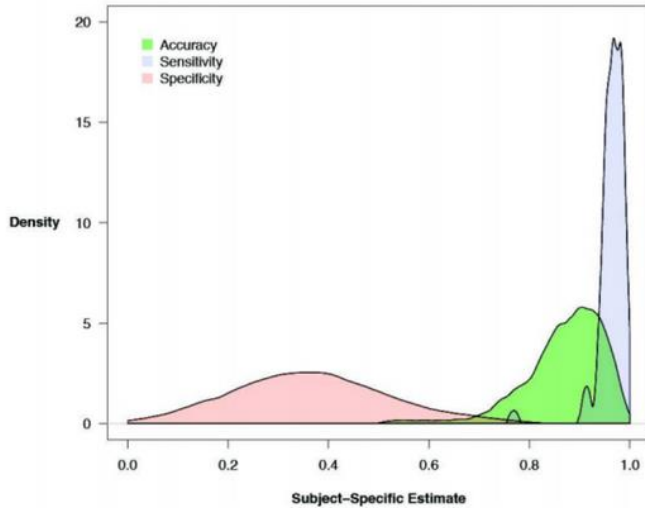
Fuente:

Bland-altman plot	Ejemplo de la gráfica Bland-Altman
Test	Prueba
Mean	Media
Mean of test 1 and test 2	Media de la prueba 1 y 2

Las gráficas de Bland-Altman también se utilizan con frecuencia para comparar metodologías. El eje X muestra la media y el eje Y muestra la diferencia entre los dos métodos. En el ejemplo de la imagen 3, la diferencia media entre las puntuaciones de dos métodos diferentes es de -6,5 unidades. Sin embargo, en los datos del gráfico de dispersión aparece una pendiente negativa. Esto significa que, a medida que la puntuación aumenta, la concordancia entre los dos métodos disminuye.

Imagen 4. Rendimiento de AW-64 y Actiwatch Spectrum





Fuente: Marino *et al.*, 2013, <https://bit.ly/3D1rP4L>.

Accuracy	Precisión
Sensitivity	Sensibilidad
Specificity	Especificidad
Density	Densidad
Subject specific estimate	Estimación específica del sujeto

El AW-64 (Minimitter, Inc, Bend, OR) y el Actiwatch Spectrum (Philips/Respironics, Murrysville, PA) fueron validados en 2013 en una muestra de 90 individuos (Marino et al., 2013). La sensibilidad y la precisión fueron altas, pero la especificidad fue comparativamente baja (imagen 4).

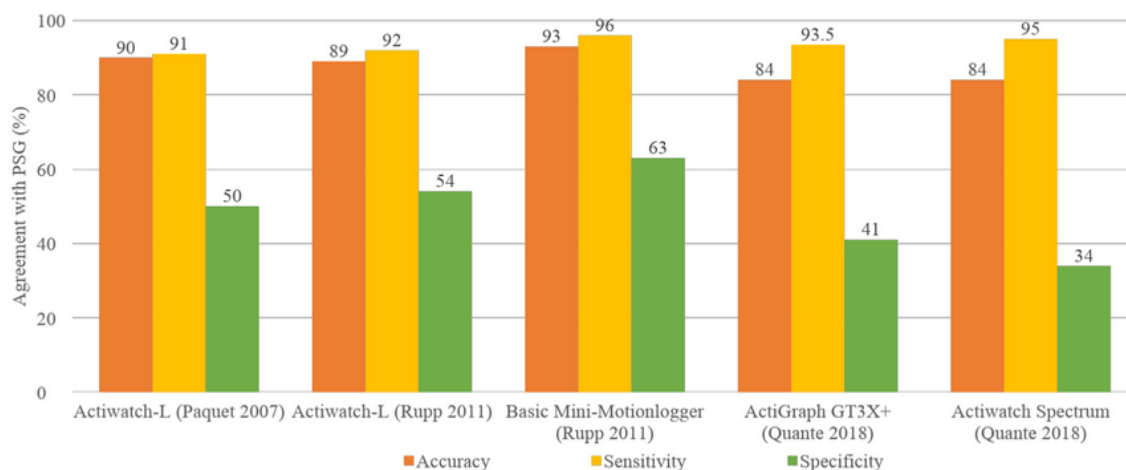
Este fue el primer estudio que abordó la importancia de la sincronización temporal entre la actigrafía y la PSG durante las grabaciones. Los investigadores compararon intervalos específicos a lo largo del período de sueño entre las grabaciones de actigrafía y PSG, alineándolos para obtener la mayor proporción de intervalos coincidentes. El estudio no encontró diferencias significativas en los resultados estratificados por género, hora del día o la presencia de insomnio; pero hubo un sesgo en la puntuación de WASO (despertares después del inicio del sueño).

Una noche de sueño con una cantidad proporcional mayor de despertares aumenta la discordancia entre la actigrafía y la PSG. Esto se debe a que la actigrafía tiende a no captar los momentos de vigilia en medio de los períodos de sueño debido a la falta general de movimiento. En esto, el gráfico de Bland Altman muestra que, a medida que aumenta el WASO, aumenta el riesgo de que los dispositivos de actigrafía informen erróneamente la vigilia nocturna como sueño. La PSG sigue capturando la vigilia correctamente, pero la actigrafía sobrepuntúa el sueño. Esto resultó en una puntuación de especificidad más baja.



Las puntuaciones de precisión, sensibilidad y especificidad para una variedad de dispositivos se muestran en la imagen 5.

Imagen 5. Rendimiento de varios dispositivos



Fuente: Lujan *et al.*, 2021

Agreement with PSG	Concordancia con PSG
Actiwatch-L	Actiwatch-L
Basic mini-motionlogger	Mini-Mtionlogger básico
actigraphy	actigrafía
Actiwatch spectrum	Actiwatch spectrum
Accuracy	Precisión
Sensitivity	Sensibilidad
Specificity	Especificidad

El registrador GENEActiv (ActivInsights Ltd., Kimbolton, Reino Unido) se comparó con el Actiwatch en un estudio de 2013 que incluyó a 15 participantes (te Lindert y Van Someren, 2013). Los investigadores aplicaron el algoritmo de Actiwatch a las puntuaciones generadas por el dispositivo GENEActiv y encontraron una concordancia más sólida entre dos dispositivos GENEActiv, o entre un dispositivo GENEActiv y un Actiwatch, que entre dos Actiwatches (tabla 3). El algoritmo utilizado fue el siguiente:

$$A_0 = [0,04E_{-(8-5)} + 0,20E_{-(4.1)} + 0,40E_{(0)} + 0,20E_{+(1-4)} + 0,04E_{+(5.8)}]$$

Donde A_0 es la actividad total reevaluada para el intervalo de interés de 15 segundos, E_0 es la actividad puntuada en el intervalo, y E_n es la actividad puntuada en los dos minutos

antes y después del intervalo puntuado. Si $A > T$, el intervalo se puntúa como vigilia. Si $A \leq T$, el intervalo se puntúa como sueño.

Tabla 5 GENEActiv versus Actiwatch

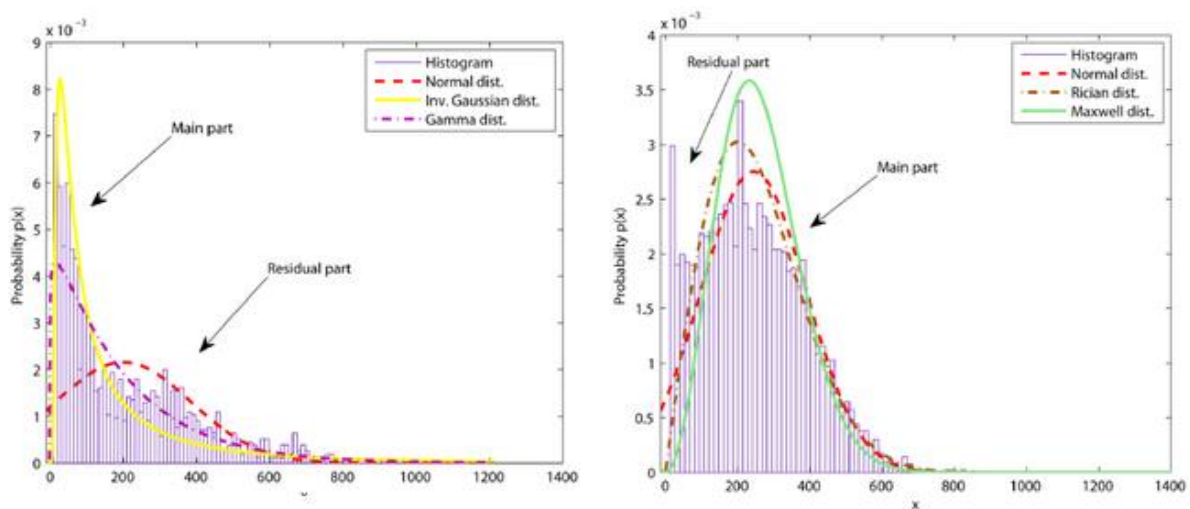
		Puntuaciones SMEM en sueño	Puntuaciones SMEM en vigilia
Puntuaciones en sueño	Actiwatch	$91,6 \pm 3,3\%$	$1,0 \pm 0,4\%$
Puntuaciones en vigilia	Actiwatch	$1,0 \pm 0,6\%$	$0,4 \pm 2,9\%$

Fuente: elaboración propia basado en Te Lindert and Van Someren, 2013.

Muchos de los dispositivos actuales en el mercado utilizan algoritmos personalizados y propietarios que no se divulgan a terceros, por ejemplo, a los investigadores que utilizan los dispositivos. Esto dificulta hacer una evaluación rigurosa. Otro problema surge cuando estos algoritmos se actualizan durante el transcurso de estudios de investigación en curso. Aunque estos algoritmos tienden a basarse en el algoritmo Cole-Kripke u otros algoritmos publicados previamente, cambios relativamente menores pueden dar lugar a diferencias significativas en las puntuaciones de sueño. Esto no solo dificulta la recopilación de datos de participantes y deportistas, si algunos de sus datos se han puntuado de manera diferente, sino que también plantea la pregunta sobre si los nuevos algoritmos anulan los estudios de validación anteriores.

Una forma de evitar algunos de estos problemas es entender matemáticamente cómo se distribuyen los datos actigráficos para predecir mejor el sueño versus la vigilia dentro de una ventana determinada. Observar los comportamientos circadianos puede ayudar a predecir la vigilia tranquila durante el día y la noche (figura 6; Adamec et al., 2010).

Imagen 6. Sueño versus vigilia



Fuente: Adamec *et al.*, 2010, p. 1.

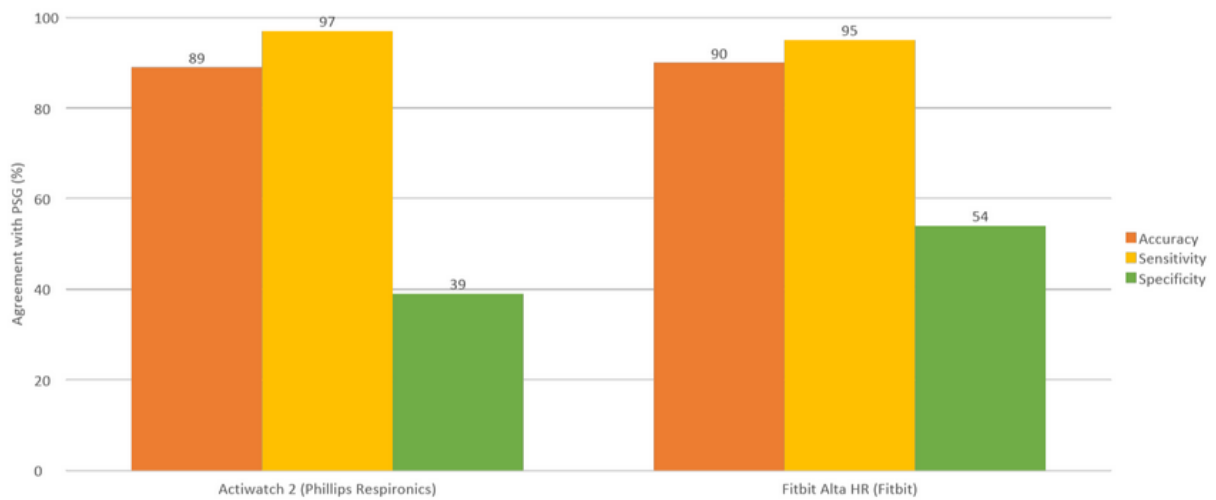
Histogram	Histograma
-----------	------------



Normal dist.	Distribución normal
Inv. Graussian dist.	Inv. Distribución gaussiana
Gamma dist.	Distribución gamma
Main part	Parte principal
Residual part	Parte residual
Probability (pix)	Probabilidad (pix)
Rician dist.	Distribución riciana
Maxwell dist.	Distribución de Maxwell

Agregar la frecuencia cardíaca como una medida adicional más allá del movimiento por sí solo puede mejorar significativamente la especificidad. En la imagen 7, el Fitbit Alta HR (Fitbit) es un 15% mejor para detectar la vigilia que el Actiwatch 2 (Philips/Respironics).

Imagen 7. Fitbit Alta HR versus Actiwatch 2



Fuente: Lujan *et al.*, 2021.

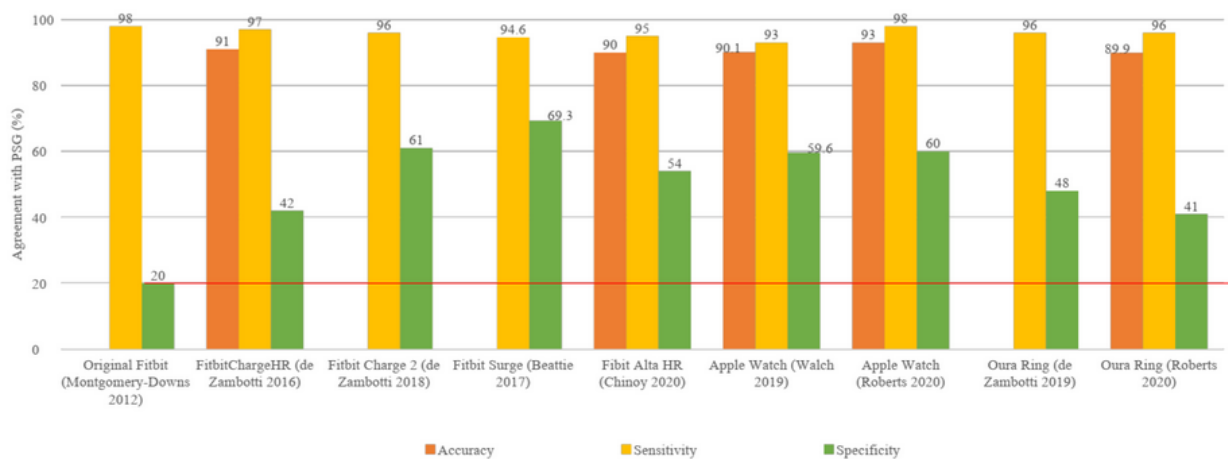
Agreement with PSG	Concordancia con PSG
Actiwatch 2 (Philips Respironics)	Actiwatch 2 (Philips Respironics)
Fitbit alta HR (fitbit)	Fitbit alta HR (fitbit)

Se puede observar una comparación de los dispositivos de próxima generación disponibles en la actualidad con el Fitbit original (que no registraba la frecuencia cardíaca) en la imagen 8.

En lo que respecta a la detección de las etapas del sueño, los datos de validación del Fitbit de 2017 mostraron una predicción general, pero con un margen de error elevado (Beattie *et al.*, 2017). Una comparación más reciente de tres dispositivos populares se puede ver en la imagen 9. Aunque hay cierta variabilidad, sigue habiendo un límite de precisión de alrededor del 90%.

Imagen 8. Comparación de varios dispositivos

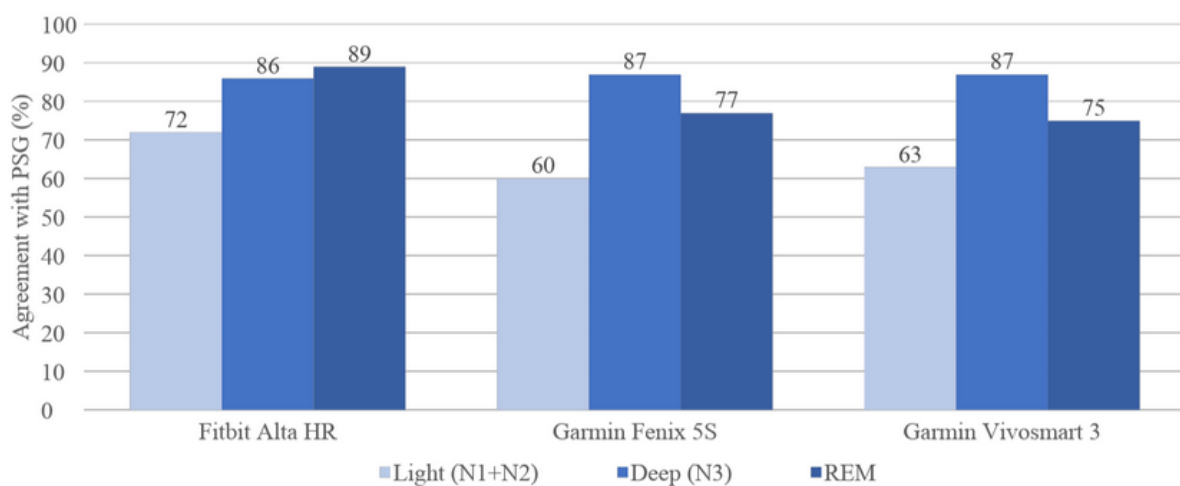




Fuente: Lujan *et al.*, 2021.

Agreement with PSG	Concordancia con PSG
Original fitbit	Fitbit original
FitbitChargHR	FitbitChargHR
Fitbit surge	Fitbit surge
Fitbit alta	Fitbit alta
Apple watch	Apple watch
Oura ring	Anillo Oura

Imagen 9. Comparación de dispositivos Fitbit versus dispositivos Garmin



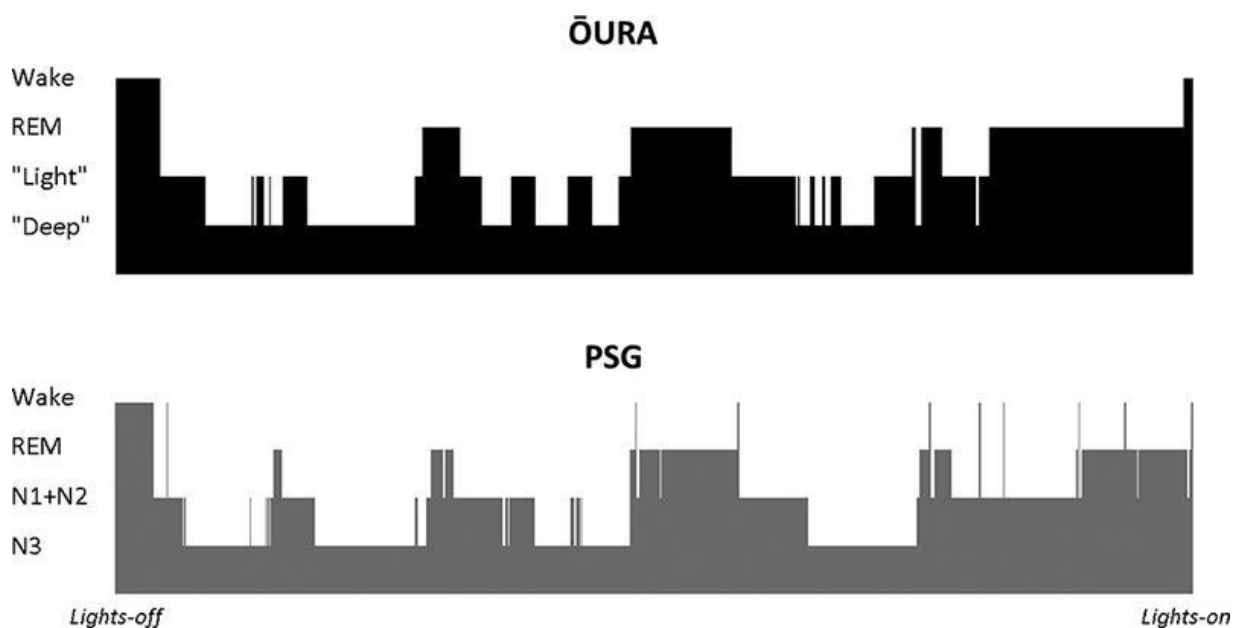
Fuente: Lujan *et al.*, 2021.

Garmin Fenix	Garmin Fenix
Garmin vivosmart	Garmin vivosmart
Light	Luz

Deep	Sueño profundo
REM	REM

La imagen 10 y la tabla 6 muestran el rendimiento del anillo Oura para detectar el sueño versus la vigilia, junto con la arquitectura del sueño, en comparación con los datos PSG. Los datos de Oura se calcularon en función de la primera versión del algoritmo de Oura en un estudio que incluyó a 41 adolescentes y adultos jóvenes, con puntajes de sensibilidad, especificidad y precisión comparables a muchos dispositivos de muñeca disponibles en el mercado.

Imagen 10. Validación de los datos de Oura



Fuente: De Zambotti *et al.*, 2019.

Oura	Oura
Wake	Vigilia
REM	REM
Light	Luz
Deep	Sueño profundo
Lights-off	Luz apagada
Lights-on	Luz encendida

Tabla 6 Validación de los datos de Oura

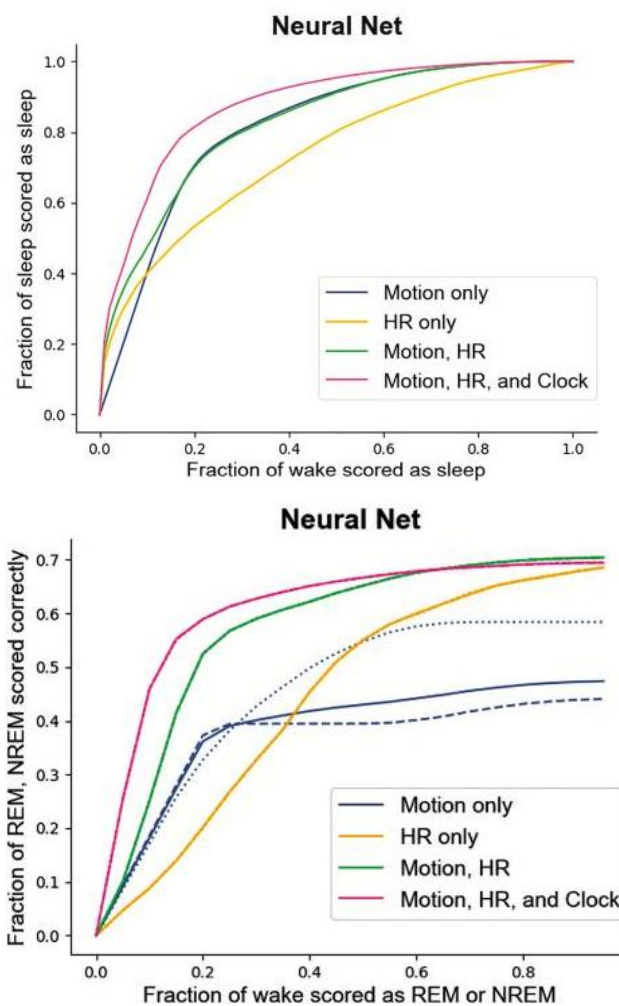
	Medias + DE	+ 95% CI del sesgo
Sensibilidad	95,5 ± 4,5	94,1-96,9
Especificidad	48.1 ± 19.1	42,0-54,1

Concordancia para N1+N2 ('sueño liviano')	64,6 ± 13,9	60,3-69,0
Concordancia para N3 ('sueño profundo')	50,9 ± 24,5	43,2-58,6
Concordancia para sueño REM	61,4 ± 22,8	54,2-68,6

Fuente: De Zambotti *et al.*, 2019.

Se muestran datos que comparan el Apple Watch (Apple Inc.) con PSG en la imagen 11. Este estudio incluyó a 39 participantes sanos que utilizaron un algoritmo diseñado por los investigadores para el estudio (Walch *et al.*, 2019). Hasta el momento de redactar esto, no existe una aplicación nativa para este dispositivo, por lo que los investigadores utilizaron los datos sin procesar. Utilizando los datos de la frecuencia cardíaca y el reloj como un indicador del ritmo circadiano, los resultados demostraron una buena estimación del sueño versus la vigilia, y una estimación moderada del sueño REM versus el sueño no REM.

Imagen 11. Comparación del Apple Watch con el PSG



Fuente: Walch *et al.*, 2019, <https://bit.ly/3SaFJpn>.

Neural Net	Red neuronal
Fraction of sleep scored as sleep	Fracción de sueño puntuada como sueño
Fraction of wake scored as sleep	Fracción de vigilia puntuada como sueño



Motion only	Solo movimiento
HR only	Solo frecuencia cardíaca
Motion HR	Movimiento y frecuencia cardíaca
Motion HR and Clock	Movimiento, frecuencia cardíaca y reloj
Fraction of REM, NREM scored correctly	Fracción de sueño REM, sueño NREM puntuado correctamente
Fraction of wake scored as REM or NREM	Fracción de vigilia puntuada como REM o NREM

Incorporar la hora del reloj en el algoritmo permitió a los investigadores ponderar de manera diferente la probabilidad de que el movimiento represente vigilia (o la quietud represente sueño) según la hora del día y en diferentes etapas del sueño. Durante la etapa 3/N3/SWS que suele ocurrir intermitentemente durante la primera mitad de la noche, el cuerpo está más relajado y es menos probable que haya movimiento. La frecuencia cardíaca también es más lenta durante esta etapa del sueño. Durante el sueño REM, que generalmente ocurre durante la segunda mitad de la noche, los músculos están prácticamente paralizados, pero los datos fisiológicos muestran una frecuencia cardíaca más similar a la que ocurre durante la vigilia. Combinar estas medidas multisensoriales puede aproximar las etapas del sueño con cierto grado de precisión, especialmente cuando se incorpora la hora del reloj en el algoritmo de puntuación.

En 2021, el Whoop fue comparado con PSG y tuvo un rendimiento razonablemente bueno en la puntuación del sueño versus la vigilia (90% de sensibilidad, 60% de especificidad, 86% de precisión). Los datos de etapas del sueño también fueron puntuados con una precisión del 60-66% (tabla 7).

Tabla 7 Validación de Whoop

Four-Stage Error Matrix for WHOOP-AUTO and PSG					
		WHOOP-AUTO			
	Stage	Wake	Light sleep	SWS	REM
PSG	Wake	60%	26%	1%	12%
	Light sleep	14%	61%	10%	15%
	SWS	6%	28%	64%	2%
	REM	6%	27%	1%	66%
	**Notes: This matrix presents the percentage of each sleep stage that WHOOP-AUTO has correctly or incorrectly classified compared to PSG. Shaded cells indicate correctly classified sleep. SWS; slow wave sleep, REM; rapid eye movement sleep.				

Fuente:



Four-stage error matrix for WHOOP-AUTO and PSG	Matriz de error de cuatro etapas para WHOOP-AUTO y PSG
PSG	PSG
Stage	Etapas
Wake	Vigilia
Light sleep	Sueño liviano
SWS	SWS
REM	REM
** Notes: This matrix presents the percentage of each sleep stage that WHOOP-AUTO has correctly or incorrectly classified compared to PSG. Shaded cells indicate correctly classified sleep. SWS; slow wave sleep, REM; rapid eye movement sleep.	**Notas: Esta matriz presenta el porcentaje de cada etapa del sueño que WHOOP-AUTO ha clasificado correctamente o incorrectamente en comparación con PSG. Las celdas sombreadas indican el sueño clasificado correctamente. SWS; sueño de ondas lentas, REM; sueño de movimientos oculares rápidos.

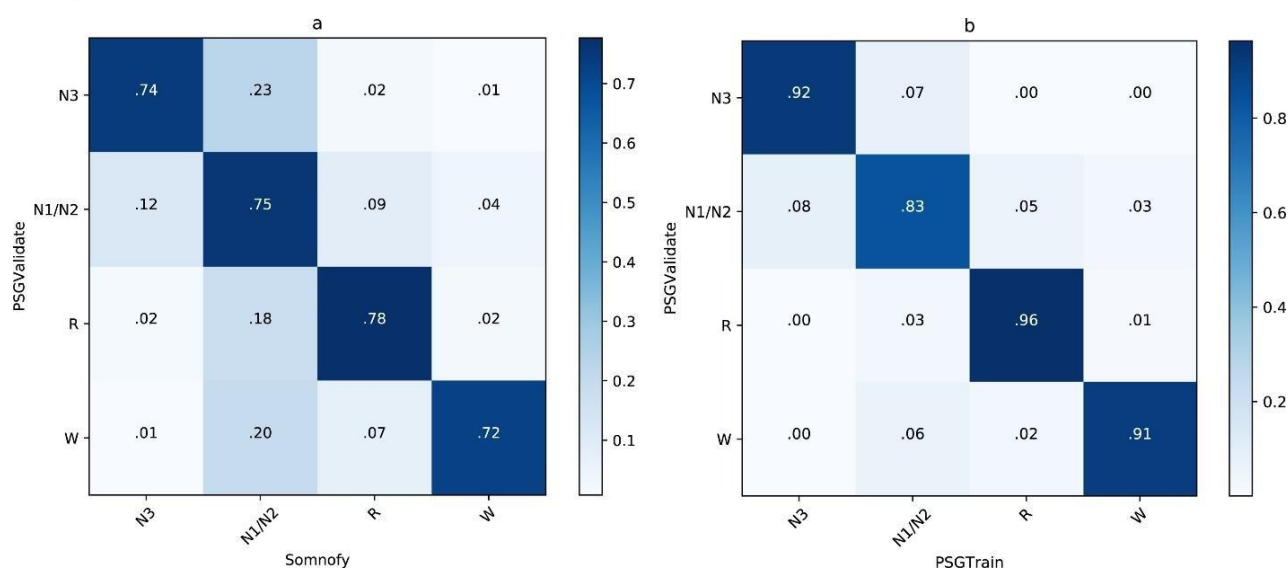
Nearables

Si bien el seguimiento del sueño generalmente se realiza mediante dispositivos ponibles, otros evalúan el sueño versus la vigilia en función de dispositivos que se mantienen cerca de la persona que duerme ("earable"). Esto incluye sensores en, sobre o cerca de la cama.

La radiofrecuencia, como el S+/SleepScore (ResMed), demostró poder detectar la respiración con una precisión moderada desde el lado de la cama, con un posible uso para el diagnóstico de la apnea del sueño. Un estudio de 2017 que evaluó su capacidad para puntuar el sueño versus la vigilia demostró una sensibilidad del 93-94%, una especificidad del 70-73% y una precisión en la clasificación del sueño del 61-62%.

El Somnofy (VitalThings) utiliza tecnología de radar para monitorear el sueño desde el lado de la cama. Los resultados de un estudio de rendimiento de 2020 se pueden ver en la tabla 8 y la imagen 12 (Toften et al., 2020)

Imagen 12. Rendimiento de Somnofy



Fuente: Toften et al., 2020, p. 59.

PSGValidate	Validación de PSG
Somnofy	Somnofy
PSGtrain	PSGtrain

Tabla 8 Rendimiento de Somnofy

	PSG	Somnofy
	Medias (DE)	Medias (DE)
Kappa de Cohen	0,82 (0,10)	0,63 (0,10)
Precisión	0,88 (0,06)	0,76 (0,07)
Sensibilidad	0,99 (0,02)	0,97 (0,03)
Especificidad	0,85 (0,11)	0,72 (0,19)

Fuente: Toften *et al.*, 2020, p. 59.

Los sensores debajo de la cama son otra forma de rastrear el sueño de manera no invasiva y son más discretos que las tecnologías utilizadas en la mesita de noche. Sin embargo, estos dispositivos tienen sus limitaciones, ya que no capturan la actividad fuera de la cama. Esto puede ser útil si la duración de interés es exclusivamente el sueño nocturno. Sin embargo, las siestas que se toman en lugares distintos a la cama, por ejemplo, no se registrarían. El rendimiento del EarlySense Live (EarlySense) en comparación con PSG se puede ver en la tabla 9 (Tal *et al.*, 2017). Las etapas del sueño se aproximan mediante cambios de presión y detección de latidos.

Tabla 9 Rendimiento de EarlySense

		Reference Values (full PSG)			
A	Piezoelectric contact-free system	Awake	REM	LS	SWS
	Awake	9,482 (80.4%)	588 (5.4%)	3,710 (9.7%)	114 (1.1%)
	REM	569 (4.8%)	5,844 (53.7%)	4,224 (11.1%)	308 (3.0%)
	LS	1,477 (12.5%)	4,190 (38.5%)	24,771 (64.9%)	4,014 (39.7%)
	SWS	265 (2.2%)	254 (2.3%)	5,471 (14.3%)	5,684 (56.2%)
		PSG Reference			
B	Piezoelectric contact-free system	Wake	Asleep		
	Wake	9,482 (80.4%)	4,412 (7.5%)		
	Asleep	2,311 (19.6%)	54,760 (92.5%)		

Contingency tables comparing values obtained with the piezoelectric contact-free system to those acquired with full polysomnography for all subjects in all three setups (85 nights).

Fuente: Tal *et al.*, 2017, p. 519.

Piezoelectric contact-free system	Sistema piezoeléctrico sin contacto
Awake	Despierto
REM	REM
LS	Sueño ligero
SWS	SWS
Piezoelectric contact-free system	Sistema piezoeléctrico sin contacto
Wake	Vigilia
Asleep	Dormido
Reference values (full PSG)	Valores de referencia (PSG completo)



PSG reference	PSG de referencia
Wake	Vigilia
Asleep	Dormido
Contingency tables comparing values obtained with the piezoelectric contact-free system to those acquired with full polysomnography for all subjects in al three setups (85 nights)	Tablas de contingencia que comparan los valores obtenidos con el sistema piezoeléctrico sin contacto con los obtenidos mediante la polisomnografía completa para todos los sujetos en los tres escenarios (85 noches).

Beddit es otro sensor que se coloca debajo de la ropa de cama sobre el colchón. Registra la frecuencia cardíaca, los ronquidos, la respiración y la presión, y puede aproximar múltiples parámetros del sueño. Los datos que comparan el dispositivo con PSG se pueden ver en las tablas 10 y 11 (Tuominen et al., 2019).

Tabla 10 Rendimiento de Beddit

	PSG			BST			P
	n	Medias (DE)	Rango	n	Medias (DE)	Rango	
TST (mins)	19	412,8 (58,1)	333.0-524.4	19	456,3 (52,0)	363.0-531.6	<0,001
SOL (mins)	17	34,1 (18,0)	5.5-70.0	17	30,9 (20,4)	8,0-86,0	NS
WASO (mins)	18	46,1 (33,0)	13,0-121,5	18	13,5 (17,3)	0,0-56,0	<0,001
SE (%)	18	83,6 (7,9)	68,0-91,8	18	90,6 (6,7)	78,2-98,2	<0,001

Fuente: Tuominen *et al.*, 2019, p. 485.

Abreviaturas: TST, tiempo total de sueño; SOL, latencia del inicio del sueño; WASO, tiempo de vigilia después del inicio del sueño; SE, eficiencia del sueño.

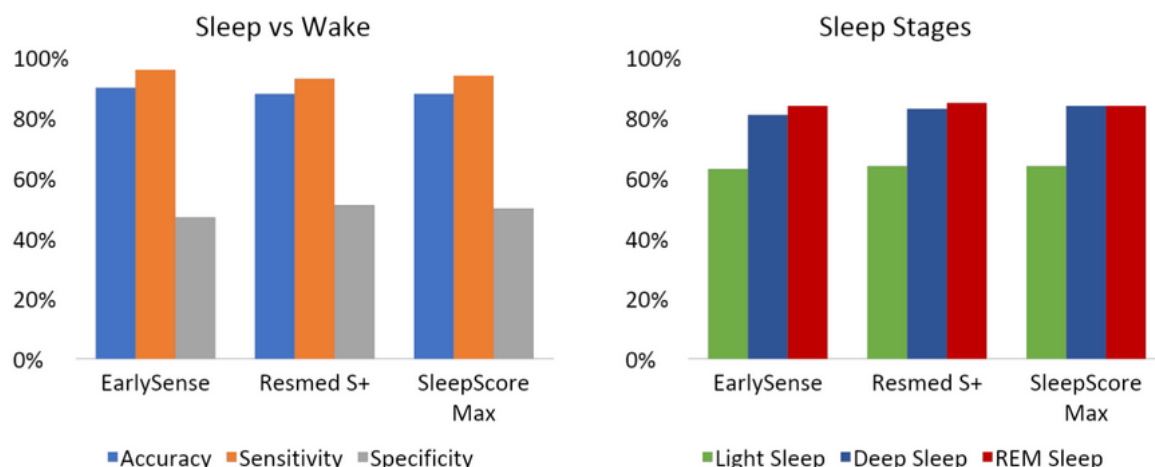
Tabla 11 Rendimiento de Beddit

		PSG			
		Vigilia	N1 (%)	N2 (%)	N3 (%)
BST	Vigilia	42,1	9,9	40,2	7,8
	N1 = Sueño ligero (%)	31,4	9,6	45,3	13,7
	N2 = Ligero (%)	15,4	8,7	49,6	28,3
	N3 = Sueño profundo (%)	5,8	4,8	51,9	37,5

Fuente: Tuominen *et al.*, 2019, p. 485.

Un estudio de 2021 comparó el rendimiento de EarlySense Live, Resmed S+ y SleepScore en términos de puntuar el sueño versus la vigilia y aproximar las etapas del sueño en comparación con PSG (Chinoy et al., 2021). Como se ve en la imagen 13, los datos son relativamente similares a muchos de los dispositivos ponibles populares que existen actualmente en el mercado.

Imagen 13. Rendimiento de EarlySense, Resmed S+ y Sleepscore



Fuente:

Sleep vs Wake	Sueño versus vigilia
EarlySense	EarlySense
Resmed S +	Resmed S +
SleepScore Max	SleepScore Max
Accuracy	Precisión
Sensitivity	Sensibilidad
Specificity	Especificidad
Sleep stages	Etapas del sueño
earlySense	earlySense
Resmed S+	Resmed S +
SleepScore Max	SleepScore Max
Light sleep	Sueño liviano
Deep sleep	Sueño profundo
REM Sleep	Sueño REM

Las bandas ponibles de EEG son una categoría de dispositivos menos discutida que aproxima el sueño al registrar la actividad cerebral además de la acelerometría. El Sleep Profiler (Advanced Brain Monitoring) registra EEG, EOG y EMG, junto con la posición y movimiento de la cabeza. Un estudio de validación realizado por Finan y sus colegas en 2016 mostró una concordancia del 92% en la vigilia, un 89% en el sueño N1, un 78% en el sueño N2, un 88% en el sueño N3 y un 85% en el sueño REM. Los datos de las etapas del sueño tienden a ser notablemente mejores en estos dispositivos en comparación con los que se usan en las muñecas o los dedos.

El dispositivo Dreem headband (Dreem) es otro que ha ganado popularidad en los últimos años. Arnal y sus colegas (2020) evaluaron el rendimiento del dispositivo Dreem headband en comparación con PSG. Los puntajes de concordancia se pueden ver en la imagen 14.

Otra medida que se está incorporando en los dispositivos ponibles es la saturación de oxígeno. Disminuciones intermitentes por debajo del nivel de referencia típico, que se encuentran entre el 95% y el 98%, pueden indicar apnea del sueño (Ryan et al., 1989).

Imagen 14. Matriz de confusión de Dreem headband versus PSG

	Wake	N1	N2	N3	REM
Wake	74.0% (2400)	12.4% (341)	8.1% (215)	1.1% (26)	4.4% (109)
N1	17.8% (278)	47.7% (764)	20.6% (344)	1.2% (17)	12.7% (203)
N2	1.4% (175)	3.2% (404)	82.9% (10100)	6.8% (822)	5.6% (715)
N3	0.1% (5)	0.1% (2)	16.4% (377)	82.6% (2660)	0.8% (27)
REM	3.1% (150)	3.0% (144)	9.4% (363)	0.0% (0)	84.5% (4021)
	DH				

Fuente: Arnal *et al.*, 2020.

concensus	consenso
-----------	----------

La conductancia cutánea y la temperatura también se pueden utilizar para aproximar mejor el sueño. La conductancia cutánea disminuye durante el sueño, con el punto más bajo que suele ocurrir durante el sueño de ondas lentas (SWS). La temperatura corporal central sigue un patrón similar (Barrett *et al.*, 1993). Cuando se combinan con datos de movimiento y frecuencia cardíaca, estos parámetros pueden proporcionar una mayor comprensión de la fisiología del sueño.

Otros dispositivos ponibles se comercializan como capaces de evaluar marcadores inflamatorios y cambios hormonales durante la noche. La interleucina-6 (CRP), por ejemplo, es una citocina que exhibe un comportamiento rítmico y depende, en parte, del sueño. La proteína C reactiva (CRP) es otra molécula que depende del sueño y que se eleva en la privación del sueño (Haack *et al.*, 2007).

El cortisol es una hormona que está intrínsecamente ligada a los ritmos circadianos innatos y exhibe un pico durante la vigilia matutina y su punto más bajo durante las etapas más profundas del sueño. Los dispositivos que pueden incorporar parámetros biológicos como este en la puntuación del sueño pueden ser más precisos. Sin embargo, inevitablemente serán menos discretos, lo que puede ser menos preferible para ciertos estudios o preferencias de los consumidores.



Estándares para la validación

Un artículo de 2020 dirigido a fabricantes de nuevas tecnologías para el sueño fue creado por investigadores del sueño como una "hoja de ruta" de estándares que incluye información sobre diversas métricas que deben considerarse al evaluar el rendimiento de un nuevo dispositivo (Depner et al., 2020). Se discuten los diversos indicadores para medir el sueño, métricas relevantes del ritmo circadiano y el sueño, y el proceso correcto para evaluar los nuevos dispositivos en comparación con PSG. Se destaca la importancia de tener acceso a datos de intervalo en intervalo para realizar análisis rigurosos, en lugar de tener solo estadísticas resumidas.

Otras consideraciones importantes incluyen el algoritmo que se utilizó para predecir el sueño frente a la vigilia o los datos de las etapas del sueño. Hasta el momento, no existe un consenso sobre lo que se considera una concordancia aceptable con PSG. Dado que la mayoría de los dispositivos tienen dificultades para detectar la vigilia tranquila y, por lo tanto, la especificidad suele ser baja. Los dispositivos a menudo son comparables entre sí, pero es posible que no lo sean de manera sólida con PSG. Si los investigadores, clínicos y consumidores están bien informados sobre las limitaciones del dispositivo y ponen las grabaciones en un contexto más amplio, pueden desempeñar un papel valioso, pero a menudo no capturan datos con el nivel de precisión que anuncian en sus promociones.

Otros parámetros de rendimiento incluyen la tasa de falla del dispositivo, ya que algunos pueden fallar con más frecuencia que otros. También es importante destacar quiénes fueron evaluados utilizando el dispositivo en un estudio dado. La diversidad en edad, peso, tono de piel y otros parámetros puede afectar los resultados. Además, ¿se realizó el estudio con dispositivos usados en un laboratorio o en casa? Los estudios realizados en un laboratorio pueden estar más controlados, pero los resultados de los estudios en casa son más generalizables.

Algunos estudios han hecho que los participantes usen múltiples dispositivos al mismo tiempo. Esto puede ser adecuado para dispositivos dirigidos a diferentes partes del cuerpo (como una vincha, un anillo y un dispositivo para la muñeca), pero será menos efectivo para múltiples anillos o múltiples dispositivos para la muñeca. En última instancia, los resultados de los dispositivos usados en la parte habitual serán preferenciales.

Históricamente, la palabra "validación" se ha utilizado en estudios de rendimiento que evalúan la tecnología ponible de sueño en comparación con PSG. Un comentario de 2021 de Cathy A. Goldstein y Christopher Depner, dos destacados investigadores del sueño, sugiere cambiar esta expresión a "rendimiento en contexto". Los criterios muy específicos que a menudo se emplean en estos estudios pueden no permitir que los resultados sean generalizables a la población en general. En lugar de utilizar la palabra "validación", que



sugiere que se requiere un solo evento para considerar que los dispositivos ponibles son aceptables para uso generalizado, la "evaluación del rendimiento" continúa investigando la eficacia bajo diversas condiciones y poblaciones. Sin embargo, se necesita una velocidad de publicación más rápida para ayudar a cerrar la brecha entre la rapidez de la innovación y el lento ritmo del proceso de revisión por pares de los artículos científicos.

Otra limitación es la resolución actual con la que los algoritmos clasifican el sueño frente a la vigilia. Los algoritmos clásicos generalmente funcionan como filtros paso bajo, con valores por encima o por debajo de un punto de corte específico clasificados como sueño o vigilia. Al agregar intervalos en un filtro de respuesta impulso finito (FIR, por sus siglas en inglés), esto puede mejorar la capacidad de estos dispositivos para calificar con precisión el comportamiento de sueño-vigilia (Bieganski et al., 2021).

Referencias

- Adamec, O., Domingues, A., Paiva, T., Sanches, J. M.** (2010). Statistical characterisation of actigraphy data during sleep and wakefulness states. *IEEE Engineering in Medicine and Biology Society*. 2342–2345.
- Arnal, P., Thorey, V., Debellemanni, E., Ballard, M., Hernandez, A., Guillot, A., & Jourde, H., Harris, M., Guillard, M., Beers, P., Chennaoui, M., & Sauvet, F.** (2020). The Dreem Headband compared to Polysomnography for EEG Signal Acquisition and Sleep Staging. *Sleep*. <http://dx.doi.org/10.1093/sleep/zsaa097>.
- Barrett, J., Lack, L., & Morris, M.** (1993). The Sleep-Evoked Decrease of Body Temperature. *Sleep*, 16(2), 93–9. <http://dx.doi.org/10.1093/sleep/16.2.93>.
- Beattie, Z., Pantelopoulos, A., Ghoreyshi, A., & Oyang, Y.** (2017). Estimation of sleep stages using cardiac and accelerometer data from a wrist-worn device. *Sleep*. <http://dx.doi.org/10.1093/sleepj/zsx050.067>.
- Biegański, P., Stróż, A., Dovgialo, M., Duszyk-Bogorodzka, A., & Durka, P.** (2021). On the Unification of Common Actigraphic Data Scoring Algorithms. *Sensors (Basel, Switzerland)*, 21(18), 6313. <https://doi.org/10.3390/s21186313>.
- Chinoy, E. D., Cuellar, J. A., Huwa, K. E., Jameson, J. T., Watson, C. H., Bessman, S. C., Hirsch, D. A., Cooper, A. D., Drummond, S., & Markwald, R. R.** (2021). Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep*, 44(5), zsaa291. <https://doi.org/10.1093/sleep/zsaa291>.
- Cole, R. J., Kripke, D. F., Gruen, W., & Mullaney, D. J.** (1992). Technical Note Automatic Sleep/Wake Identification From Wrist Activity. *Sleep*, 15(5), 461–9. <http://dx.doi.org/10.1093/sleep/15.5.461>.
- Depner, C., Cogswell, D., Bisesi, P., Markwald, R., Cruickshank-Quinn, C., Quinn, K., Melanson, E., Reisdorph, N., & Jr, K.** (2020). Developing preliminary blood



metabolomics-based biomarkers of insufficient sleep in humans. *Sleep*, 43. <http://dx.doi.org/10.1093/sleep/zsz321>.

- Finan, P. H., Quartana, P. J., Remeniuk, B., Garland, E. L., Rhudy, J. L., Hand, M., Irwin, M. R., & Smith, M. T.** (2016). Partial sleep deprivation attenuates the positive affective system: Effects across multiple measurement modalities. *Sleep*, 40. <https://doi.org/10.1093%2Fsleep%2Fzsw017>.
- Haack, M., Sanchez, E., & Mullington, J. M.** (2007). Elevated inflammatory markers in response to prolonged sleep restriction are associated with increased pain experience in healthy volunteers. *Sleep*, 30, 1145.
- Kripke, D. F., Mullaney, D. J., Messin, S., & Wyborney, V. G.** (1978). Wrist actigraphic measures of sleep and rhythms. *Electroencephalography and Clinical Neurophysiology*, 44(5), 674-6. [https://doi.org/10.1016/0013-4694\(78\)90133-5](https://doi.org/10.1016/0013-4694(78)90133-5).
- Kripke, D. F., Hahn, E. K., Prorock Grizas, A., & Wadiak, K. H.** (2010). Wrist actigraphic scoring for sleep laboratory patients: Algorithm development. *Journal of Sleep Research*, 19(4), 612-9. <http://dx.doi.org/10.1111/j.1365-2869.2010.00835.x>.
- Kupfer, D. J., Himmelhoch, J. M., Swartzburg, M., Anderson, C., Byck, R., Detre, T. P.** (1972). Hypersomnia in manic-depressive disease: A preliminary report. *Diseases of the Nervous System*, 33(11), 720-724.
- Leproult, R., Copinschi, G., Buxton, O., & Van Cauter, E.** (1997). Sleep loss results in an elevation of cortisol levels the next evening. *Sleep*, 20(10), 865-870.
- Marino, M., Li, Y., Rueschman, M. N., Winkelman, J. W., Ellenbogen, J. M., Solet, J. M., Dulin, H., Berkman, L. F., & Buxton, O. M.** (2013). Measuring Sleep: Accuracy, Sensitivity, and Specificity of Wrist Actigraphy Compared to Polysomnography. *Sleep*, 36(11), 1747-1755. <https://doi.org/10.5665%2Fsleep.3142>.
- Morgenthaler, T., Kramer, M., Alessi, C., Friedman, L., Boehlecke, B., Brown, T. Coleman, J., Kapur, V., Lee Chiong, T., Owens, J., Pancer, J., & Swick, T.** (2006). Practice Parameters for the Psychological and Behavioural Treatment of Insomnia: An Update. An American Academy of Sleep Medicine Report. *Sleep*, 29, 1415-1419.
- Mullaney, D. J., Kripke, D. F., & Messin, S.** (1980). Wrist-Actigraphic Estimation of Sleep Time. *Sleep*, 3(1), 83-92.
- Okudaira, N., Fukuda, H., Nishihara, K., Ohtani, K., Endo, S., & Torii, S.** (1983). Sleep apnoea and nocturnal myoclonus in elderly persons in Vilcabomba, Ecuador. *Journal of Gerontology*, (38), 436-438.
- Ryan, T., Mlynczac, S., Ericson, T., Paul Man, S. F., & Godfrey, C. W.** (1989). Oxygen Consumption During Sleep: Influence of Sleep Stage and Time of Night. *Sleep*, 12(3), 201-210.
- Sadeh, A., Lavie, P., Scher, A., Tirosh, E., & Epstein, R.** (1991). Actigraphic home-monitoring of sleep-disturbed and control infants and young children: A new method for paediatric assessment of sleepwake patterns. *Pediatrics*, 87,494-499.



- Sadeh, A., Hauri, P. J., Kripke, D. F., Lavie, P.** (2005). The role of actigraphy in the evaluation of sleep disorders. *Sleep*, 18, 288–302.
- Tal, A., Shinar, Z., Shaki, D., Codish, S., & Goldbart, A.** (2017). Validation of Contact-Free Sleep Monitoring Device with Comparison to Polysomnography. *Journal of Clinical Sleep Medicine*, 13(3), 517-522. <https://doi.org/10.5664/jcsm.6514>.
- Te Lindert, B. H. W., & Van Someren, E. J. W.** (2013). Sleep estimates using microelectromechanical systems (MEMS). *Sleep*, 36(5), 781-9.
- Toften, S., Pallesen, S., Hrozanova, M., Moen, F., Grønli, J.** (2020). Validation of sleep stage classification using non-contact radar technology and machine learning (Somnofy®). *Sleep Medicine*, 75, 54-61.
- Tuominen, J., Peltola, K., Saaresranta, T. & Valli, K.** (2019). Sleep Parameter Assessment Accuracy of a Consumer Home Sleep Monitoring Ballistocardiograph Beddit Sleep Tracker: A Validation Study. *Journal of Clinical Sleep Medicine*, 15, 483-487. <http://dx.doi.org/10.5664/jcsm.7682>.
- Walch, O., Huang, Y., Forger, D., & Goldstein, C.** (2019). Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. *Sleep*. <https://doi.org/10.1093%2Fsleep%2Fzsz180>.

