INTRODUCTION

In search of novel biomarkers and treatment targets for age-related pathologies, sleep has recently attracted much attention. Disrupted sleep represents one of the earliest symptoms of Alzheimer’s disease (Lim, Kowgier, Yu, Buchman, & Bennett, 2013; Lucey et al., 2019), but might also potentiate, accelerate, or even cause cognitive pathology in old age (Shokri-Kojori et al., 2018; Vaou, Lin, Branson, & Auerbach, 2018). Because of the assumed bidirectional relationship between sleep and Alzheimer’s pathology (Ju, Lucey, & Holtzman, 2014; Mander, Winer, Jagust, & Walker, 2016; Noble & Spires-Jones, 2019; Vaou et al., 2018; Winer et al., 2019), studying the interrelations between sleep and neurocognitive aging has become increasingly popular.

Even in the absence of any diagnostically identifiable pathology, healthy aging entails fundamental yet very characteristic changes in sleep (Mander, Winer, & Walker, 2017; Ohayon, Carskadon, Guilleminault, & Vitiello, 2004; Vitiello, 2006). Overall, older adults’ sleep is shifted earlier in time (including earlier bed and wake times), gets lighter, more fragile, and shorter (Dijk, Duffy, & Czeisler, 2000; Monk, 2005). Sleep alterations are evident in the global distribution and succession of sleep stages (i.e., the macrostructure of sleep; Feinberg, 1974; Ohayon et al., 2004), but also involve a changed signature of electrical oscillations that define and characterize different sleep states (i.e., the microstructure of...
Box 1 Sleep electrophysiology at a glance

In most animal species, three different and discrete states of vigilance can be differentiated: wakefulness, REM, and NREM sleep (Vassalli & Dijk, 2009). REM and NREM sleep can be clearly distinguished based on electrophysiological characteristics (Aserinsky & Kleitman, 1953; Dement & Wolpert, 1958; Feinberg & Evarts, 1969; Loomis, Harvey, & Hobart, 1935, 1962). REM sleep is marked by the occurrence of phasic irregular and rapid eye movements, muscle atony, and desynchronized wake-like electroencephalographic (EEG) activity. NREM sleep, by contrast, is characterized by synchronous, low-frequency, high-amplitude EEG oscillations (Iber, Ancoli-Israel, Chesson, & Quan, 2007). NREM sleep can further be divided into the three substages, N1, N2, and N3 (also SWS; Iber et al., 2007; Rechtschaffen & Kales, 1968). These NREM substages form a continuum of increasing arousal threshold and sleep depth (Carskadon & Dement, 2011). At the same time, they serve as a proxy for the dominance of different sleep-specific oscillatory brain signals: Lighter NREM sleep (stage 2 or N2 sleep) is defined by the occurrence of discrete sleep spindles (i.e., rhythmically waxing and waning oscillatory events with a typical frequency of 11–16 Hz), and K-complexes (i.e., phasic negative high-amplitude EEG deflections that are followed by a positive EEG component). With deeper sleep, the number and power of EEG slow waves increases. This predominance of slow EEG waves defines the presence of SWS (stage 3 or N3 sleep).

sleep; cf. Box 1; Carrier et al., 2011; Crowley, Trinder, Kim, Carrington, & Colrain, 2002). In particular, older adults’ sleep is characterized by a drastic reduction of deep non-rapid eye movement (NREM) sleep, so-called slow-wave sleep (SWS), that is accompanied by an increase of light NREM sleep (Carrier et al., 2011; Danker-Hopfe et al., 2005; Ohayon et al., 2004; Redline et al., 2004). In contrast to NREM sleep, age-related changes in the amount of rapid eye movement (REM) sleep are typically more subtle (Redline et al., 2004; Scullin & Gao, 2018), although a pronounced reduction of REMs within this sleep stage has been reported (Darchia, Campbell, & Feinberg, 2003). Moreover, with advancing age, high-amplitude slow oscillations (<1 Hz), slow delta waves (1–4 Hz), and discrete sleep spindles (11–16 Hz) defining NREM sleep appear less often and with reduced amplitudes and altered topography (Crowley et al., 2002; Dubé et al., 2015; Fogel et al., 2012; Landolt & Borbély, 2001; Landolt, Dijk, Achermann, & Borbély, 1996; Martin et al., 2013). Diminished homeostatic sleep pressure, circadian shifts, dysregulation of neurotransmitters, and structural brain alterations may contribute to these pronounced changes in sleep during aging (for more details see Dijk et al., 2000; Mander et al., 2017; Monk, 2005; Skeldon, Derks, & Dijk, 2016; Zhong et al., 2019).

Standard criteria to describe these age-related alterations in sleep physiology are only rarely tested and validated within heterogeneous samples that include older adults (but see e.g., Ujma et al., 2015; Ujma, Simor, Steiger, Dresler, & Bódízs, 2019; Warby et al., 2014). Therefore, in the following, we aim to demonstrate the possible benefits and limitations of applying established indicators of sleep physiology in the context of aging research. We aim to raise awareness for the complexity of aging that entails widespread structural and functional brain changes that affect sleep in multiple ways (Dubé et al., 2015; Fogel et al., 2012, 2017; Muehlroth et al., 2019b). To identify meaningful biomarkers that can explain or even predict cognitive decline in old age, we need to promote the use of sensitive and age-adjusted methodology that captures within-person age-dependent changes in the physiology of sleep, rather than age-inflated interindividual differences (Lindenberger, von Oertzen, Ghisletta, & Hertzog, 2011).

2 | PREAMBLE: HOW DO WE DEFINE “OLD AGE” AND “AGING”? As an inevitable prerequisite, the identification of changes in sleep physiology during aging requires an appropriate and transparent rationale on how research should delineate and investigate “old age” and “aging.” The literature in (cognitive) neuroscience is dominated by cross-sectional group comparisons of “younger” (~20–30 years) and “older” adults (~60–80 years) (Browning & Spilich, 1981; Hedden & Gabrieli, 2004). This is grounded in the notion that the brain’s structure and function has already undergone profound alterations when humans reach their 60s (e.g., Li et al., 2004; Ohayon et al., 2004; Raz et al., 2005; Rönnlund, Nyberg, Bäckman, & Nilsson, 2005; Ziegler et al., 2012). Beyond stressing the importance of a general agreement on the definition of “old age” to ensure the comparability of research on aging, we need to highlight two challenges that accompany such a definition: First, the chronological age of individuals can deviate from their biological age. Second, age differences derived from such group comparisons cannot capture aging in terms of age-related change.

First and foremost, one should bear in mind that chronological age, that is, the time that has passed since birth, is...
a variable that does not induce or explain any development-
tal change in itself (Wohlwill, 1970). At best, it can serve as a proxy for an individual's expected functional capacity that relies on true mechanistic alterations taking place as hu-
mens age (Li & Schmiedek, 2002; MacDonald, DeCarlo, & Dixon, 2011). The conceptual and methodological pitfalls of using chronological age as an index of development or aging have been discussed elsewhere in detail (e.g., Cole & Franke, 2017; Cole, Marioni, Harris, & Deary, 2019; Lindenberger & Pötter, 1998; MacDonald et al., 2011; Wohlwill, 1970). Due to great interindividual variation in the manifestation of aging effects (Cabeza et al., 2018; Habib, Nyberg, & Nilsson, 2007; Lindenberger, 2014), particularly in old age, chrono-
logical age can deviate greatly from an individual's biological age (Cole & Franke, 2017; Cole et al., 2019; MacDonald et al., 2011; Steffener et al., 2016). Importantly, physiological brain age, as for instance derived from estimates of struc-
tural brain integrity, could represent a better and more infor-
mative predictor of differences in aging, individual (brain) health, functional capacity, and mortality (Burzynska et al., 2010; Cole & Franke, 2017; Cole et al., 2017, 2018; Franke, Ziegler, Klöppel, & Gaser, 2010). Notably, also brain age de-
rivered from the sleep electroencephalogram (EEG) has lately shown its potential to serve as a sensitive index of aging and pathology (Sun et al., 2019). In the course of this article, we will suggest that the sensitivity of sleep physiology as a po-
tential biomarker of brain age, aging, and age-related disease is closely tied to a fair and adapted investigation and evaluation of sleep in aged individuals.

Strictly speaking, the term aging describes a process of change, taking place as organisms grow older. Cross-sectional age comparisons are not appropriate to inform us about these changes (Hertzog & Nesselroade, 2003; Hofer & Sliwinski, 2001; Li & Schmiedek, 2002; Lindenberger et al., 2011; Lindenberger & Pötter, 1998; Overton, 2010; Raz & Lindenberger, 2011). When sampling age-heterogeneous groups cross-sectionally, essential confounds such as co-
hort effects or sampling bias may covary with age and thus prevent or bias conclusions on intraindividual dynamics of change (Hertzog & Nesselroade, 2003; Hofer & Sliwinski, 2001; Li & Schmiedek, 2002; Lindenberger et al., 2011; Rönnlund et al., 2005; Wohlwill, 1970). For convenience, however, we will use the phrase aging independently of the respective study's sampling design in this article. The lack of longitudinal studies on sleep in adulthood does not allow for the identification of “real” age-related changes of sleep physiology. Hence, we emphasize that longi-
tudinal study designs are essential to study aging and to identify the lead-lag relations between age-related alter-
ations in brain structure and function (Hofer & Sliwinski, 2001; Li & Schmiedek, 2002; Lindenberger, 2014; Raz & Lindenberger, 2011).

3 | CHALLENGE 1: AMBIGUOUS SLEEP STAGE DEFINITIONS ACROSS AGE GROUPS

Standardized visual scoring of sleep stages is regarded as the “gold-standard” of sleep research. Defined sleep stages are thereby considered sensitive indicators of specific physiological events. However, current sleep stage definitions may not capture the same physiological processes across age groups. We suggest that genuine sleep analysis—in line with similar views from animal research—should consider the continuous nature of NREM sleep. A focus on the presence of predefined electrophysiological events, like the occurrence or coupling of slow oscillations and sleep spindles, during both stage 2 sleep and SWS is required.

3.1 | Pitfalls of visually scoring SWS

In human research, the classification of sleep stages (cf. Box 1) is based on widely accepted standardized rules that are used to score sleep by means of EEG activity, muscle tone, and eye movements (Iber et al., 2007; Rechtschaffen & Kales, 1968). SWS, as defined by the American Academy for Sleep Medicine (AASM; Iber et al., 2007), requires the presence of 0.5–2 Hz waves with a minimum amplitude of 75 μV covering more than 20% of a sleep time segment. Already in 1982, Webb noted that the frequently reported striking age-related reduction in SWS likely results from the use of a fixed amplitude threshold to define slow waves (Webb, 1982; Webb & Dreblow, 1982; cf. Figure 1a, Table 1). As also acknowledged by the AASM (Silber et al., 2007), reliably measuring SWS using an amplitude criterion below 75 μV is possible. Still, the current version of the manual for sleep scoring that defines the prevailing gold standard, relies on a minimum amplitude criterion of 75 μV. Crucially, the age-
related reduction in slow wave (<4 Hz) and slow oscillation (<1 Hz) amplitudes is one of the most consistent findings in age-comparative sleep studies (Carrier et al., 2011; Dubé et al., 2015; Mander et al., 2017; Muehlroth et al., 2019b; Ujma et al., 2019). Typically, SWS is easily identifiable in younger adults because slow waves exceed the required am-
plitude threshold. By contrast, the identification of true SWS in older adults is challenging: Due to the reduction of slow wave amplitudes below 75 μV (cf. Figure 1b,c, Table 2), time segments can often not be scored as SWS but are rather as-
signed to stage 2 sleep—although all other characteristics of the sleep epoch suggest the presence of SWS (i.e., the pre-
dominance of low-frequency EEG waves but absence of eye movements along with reduced muscle tone; cf. Figure 2).

Re-analyzing our own previously published data of 29 healthy younger (range: 19–28 years, M age = 23.55 years,
We demonstrate the effects of fixed amplitude thresholds for visual sleep scoring in age-comparative studies. Participants' nighttime sleep was monitored using ambulatory polysomnography (PSG) before and after completing an associative learning task (cf. Supporting Information for details on data and methods; Muehlroth et al., 2019b, 2019a). Subjective sleep quality was assessed using the Pittsburgh Sleep Quality Index (PSQI; Buysse, Reynolds, Monk, Berman, & Kupfer, 1989) and comparable in both age groups (Z = 1.03, p = .305, Md_{YA} = 4.5 [3.0; 6.25], Md_{OA} = 4.0 [3.0; 5.0]). We show that, overall, in our data set the extent of age-related reductions in visually scored SWS matched the increase in stage 2 sleep in older adults (d_{stage 2} = −2.3 [−2.96; −1.63]; d_{SWS} = 2.18).

### TABLE 1 Age differences in sleep architecture

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Younger Adults Md [1st qrt; 3rd qrt]</th>
<th>Older adults Md [1st qrt; 3rd qrt]</th>
<th>Z</th>
<th>p</th>
<th>Cohen’s d [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>TST (min)</td>
<td>456.50 [430.25; 496.75]</td>
<td>425.50 [385.06; 459.88]</td>
<td>2.48</td>
<td>.013</td>
<td>0.62 [0.1; 1.15]</td>
</tr>
<tr>
<td>WASO (min)</td>
<td>6.75 [3.00; 11.50]</td>
<td>21.25 [12.06; 49.81]</td>
<td>−4.23</td>
<td>&lt;.001</td>
<td>−1 [−1.54; −0.45]</td>
</tr>
<tr>
<td>Stage 1 (%)</td>
<td>4.46 [2.84; 4.82]</td>
<td>7.39 [4.71; 9.90]</td>
<td>−3.99</td>
<td>&lt;.001</td>
<td>−1.02 [−1.57; −0.48]</td>
</tr>
<tr>
<td>Stage 2 (%)</td>
<td>51.68 [46.29; 58.10]</td>
<td>65.62 [61.71; 69.14]</td>
<td>−6.23</td>
<td>&lt;.001</td>
<td>−2.95 [−2.96; −1.63]</td>
</tr>
<tr>
<td>SWS (%)</td>
<td>20.41 [16.01; 27.83]</td>
<td>2.30 [0.16; 11.15]</td>
<td>6.03</td>
<td>&lt;.001</td>
<td>2.18 [1.53; 2.83]</td>
</tr>
<tr>
<td>REM (%)</td>
<td>22.91 [20.22; 26.07]</td>
<td>22.19 [18.22; 24.00]</td>
<td>1.46</td>
<td>.144</td>
<td>0.31 [−0.20; 0.83]</td>
</tr>
</tbody>
</table>

Note: Z and p values were derived from nonparametric Mann–Whitney U tests comparing sleep parameters between younger and older adults. Sleep stage percentages are calculated as proportions of TST.

Abbreviations: CI, confidence interval; qrt, quartile; REM, rapid eye movement sleep; SWS, slow-wave sleep; TST, total sleep time; WASO, wake after sleep onset (i.e., time awake between sleep onset and final morning awakening).
TABLE 2  Individually adjusted versus fixed threshold settings for slow oscillation detection

<table>
<thead>
<tr>
<th></th>
<th>Younger Adults</th>
<th>Older adults</th>
<th>Z</th>
<th>p</th>
<th>Cohen’s d [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Md [1st qrt; 3rd qrt]</td>
<td>Md [1st qrt; 3rd qrt]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individually adjusted threshold</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>1,109.75 [1,068.75; 1,362.00]</td>
<td>917.75 [807.50; 1,133.56]</td>
<td>3.93</td>
<td>&lt;.001</td>
<td>1.09 [0.54; 1.65]</td>
</tr>
<tr>
<td>Density (events/min)</td>
<td>3.81 [3.45; 4.12]</td>
<td>3.52 [2.92; 3.93]</td>
<td>2.26</td>
<td>.024</td>
<td>0.67 [0.14; 1.2]</td>
</tr>
<tr>
<td>Frequency (Hz)</td>
<td>0.79 [0.78; 0.81]</td>
<td>0.77 [0.75; 0.79]</td>
<td>3.24</td>
<td>.001</td>
<td>0.90 [0.36; 1.44]</td>
</tr>
<tr>
<td>Amplitude (µV)</td>
<td>123.17 [106.91; 136.50]</td>
<td>71.83 [58.74; 84.53]</td>
<td>6.16</td>
<td>&lt;.001</td>
<td>2.28 [1.62; 2.94]</td>
</tr>
<tr>
<td>Fixed threshold</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>1,329.25 [1,094.00; 1,672.50]</td>
<td>350.00 [149.44; 564.81]</td>
<td>6.07</td>
<td>&lt;.001</td>
<td>2.02 [1.38; 2.65]</td>
</tr>
<tr>
<td>Density (events/min)</td>
<td>4.25 [3.24; 5.51]</td>
<td>1.12 [0.53; 2.33]</td>
<td>6.04</td>
<td>&lt;.001</td>
<td>1.93 [1.31; 2.56]</td>
</tr>
<tr>
<td>Frequency (Hz)</td>
<td>0.79 [0.78; 0.81]</td>
<td>0.76 [0.73; 0.78]</td>
<td>4.14</td>
<td>&lt;.001</td>
<td>1.16 [0.61; 1.72]</td>
</tr>
<tr>
<td>Amplitude (µV)</td>
<td>116.92 [112.46; 125.07]</td>
<td>94.85 [91.57; 102]</td>
<td>5.86</td>
<td>&lt;.001</td>
<td>2.13 [1.48; 2.77]</td>
</tr>
</tbody>
</table>

Note: Z and p values were derived from nonparametric Mann–Whitney U tests comparing slow oscillation characteristics between younger and older adults. Abbreviations: CI, confidence interval; OA, older adults; qrt, quartile; YA, younger adults.

FIGURE 2  Whole-night spectral EEG characteristics. Hypnogram (upper panels) from one younger (a; female, 24.63 years) and older (b; female, 71.55 years) adult with the corresponding whole-night multitaper spectral EEG data at electrode Cz (lower panels). Both younger and older adults show a similarly increased low-frequency power during periods of NREM sleep (some examples highlighted by blue ellipses) along with a distinct increase in power in the fast spindle frequency range (approx. 12.5–16 Hz) that disappears during REM sleep (violet ellipses). Periods of wake are characterized by an additional strong enhancement of alpha (8–12 Hz) and high-frequency power (red ellipses). Although the older adult only has a small amount of visually scored SWS, the spectrogram clearly shows periods of enhanced low frequency and spindle power—in line with the canonical SWS characteristics found in younger adults. No clear distinction between stage 2 and SWS is observed on the basis of spectral EEG characteristics. Note the differences in absolute sleep duration (x axis). OA, older adult; REM, rapid eye movement sleep; SWS, slow-wave sleep; YA, younger adult.
cepted in human sleep research, the distinction between & Battaglia, 2014; Hobson, 1968). Although generally ac-
specific oscillatory brain signals (Genzel, Kroes, Dresler, 
a II and SWS are continously with slow waves increasing in amplitude and
-2 and SWS based on a certain amplitude and time criterion (Iber et al., 2007) is not without problems (Genzel et al., 2014; Lacroix et al., 2018; Silber et al., 2007; Webb & Dreblow, 1982). Challenges include the handling of potentially biased sleep stage estimates in different age groups and the comparability of estimates between human and animal research (e.g., Genzel et al., 2014; Lacroix et al., 2018).

As discussed above, the current criteria to define the presence of SWS along with great interindividual variation in slow wave amplitudes may severely bias the assignment of a specific sleep segment to stage 2 sleep or SWS. Between the ages of 5 and 90, the proportion of NREM sleep decreases only little (Ohayon et al., 2004). The composition of visually scored NREM sleep itself, however, changes drastically across the lifespan. This might at least partly stem from the fact that time segments “equivalent” to SWS may be scored as stage 2 sleep because the amplitude criterion of 75 µV is not met and result in two major problems: First, the assignment of SWS-like time segments to stage 2 sleep in older adults blurs the boundary between the two stages. As a consequence, sleep stages can no longer be regarded as a direct reflection or proxy of underlying neural and physiological processes (Hobson, 1968). Second, due to skewed distributions and reduced variance, any correlational analysis relying on measures of SWS is biased, or even prevented, in samples of older adults—especially in samples of older adults with a large percentage of “no-SWS” participants. As a result, a clear-cut association between a given physiological state—identified via a corresponding sleep stage—and an outcome measure, like the ability to maintain memories across sleep (e.g., Diekelmann & Born, 2010; Rasch & Born, 2013), is hardly tangible in age-comparative studies. If the definitions used to identify a given sleep stage do not capture the same phenomenon (e.g., the presence of slow oscillations) to the same degree across age groups, age differences in associations with outcome measures can hardly capture the intended mechanism. For example, the observation that memory performance is more consistently associated with SWS in younger than in older adults, could be taken as evidence that slow oscillations forfeit their functionality for memory processing during aging (e.g., Baran, Mantua, & Spencer, 2016; Scullin, 2012; Sonni & Spencer, 2015). However, if slow oscillations are present in older adults but just attributed to a different sleep stage, the statistics might be correct, but the functional interpretation would be different. In this case, slow oscillations found in a different sleep stage may continue to contribute to the consolidation of memories even in old age—a matter of fact only revealed when focusing on the oscillatory phenomena themselves (e.g., Helfrich, Mander, Jagust, Knight, & Walker, 2018; Mander et al., 2013; Muehlroth et al., 2019b; Varga et al., 2016; Westerberg et al., 2012).

Taken together, it appears that due to age-related reductions in slow wave amplitudes, the fixed amplitude criterion of 75 µV for scoring SWS results in a biased estimate of SWS presence in older populations. Descriptively providing standardized estimates of sleep architecture is important to ensure the comparability and validity of studies. In studies with heterogeneous samples, though, researchers should acknowledge that such derived sleep stage estimates may be severely biased. As Rechtschaffen and Kales themselves put it in 1968 when establishing the standardized scheme for visual sleep scoring: “Even among human sub-
jects, however, there are some individuals or groups whose polygraph recordings may require further description or elaboration than that provided by the stages proposed here” (Rechtschaffen & Kales, 1968).

3.2 The need to acknowledge the continuous nature of NREM sleep

The potential bias of current sleep stage definitions across age groups has direct theoretical and methodological implications. These affect the way we should conceive and treat NREM sleep and its substages in our theories and analyses. The transition between stage 1, stage 2, and SWS happens continuously with slow waves increasing in amplitude and abundance as NREM sleep gets deeper. On the contrary, sleep spindles show a reversed pattern with predominance during lighter stage 2 NREM sleep (Prerau, Brown, Bianchi, Ellenbogen, & Purdon, 2017). Hence, stage 2 and SWS are often treated as proxies for the dominance of different sleep-specific oscillatory brain signals (Genzel, Kroes, Dresler, & Battaglia, 2014; Hobson, 1968). Although generally accepted in human sleep research, the distinction between individual differences in the amplitude of slow oscillations, with half of the older adults showing slow oscillation amplitudes below the proposed 75-µV threshold on average ($M_{OA} = 71.83$ µV; Figure 1b). For older adults, falling above or below this threshold divided the sample into individuals with and without SWS according to the standard scoring system (Figure 1c). None of the younger adults urcet an average slow oscillation amplitude threshold of 75 µV ($M_{YA} = 123.17$ µV) and, accordingly, SWS could be effectively scored in this age group (Figure 1b,c).

As discussed above, the current criteria to define the presence of SWS along with great interindividual variation in slow wave amplitudes may severely bias the assignment of a specific sleep segment to stage 2 sleep or SWS. Between the ages of 5 and 90, the proportion of NREM sleep decreases only little (Ohayon et al., 2004). The composition of visually scored NREM sleep itself, however, changes drastically across the lifespan. This might at least partly stem from the fact that time segments “equivalent” to SWS may be scored as stage 2 sleep because the amplitude criterion of 75 µV is not met and result in two major problems: First, the assignment of SWS-like time segments to stage 2 sleep in older adults blurs the boundary between the two stages. As a consequence, sleep stages can no longer be regarded as a direct reflection or proxy of underlying neural and physiological processes (Hobson, 1968). Second, due to skewed distributions and reduced variance, any correlational analysis relying on measures of SWS is biased, or even prevented, in samples of older adults—especially in samples of older adults with a large percentage of “no-SWS” participants. As a result, a clear-cut association between a given physiological state—identified via a corresponding sleep stage—and an outcome measure, like the ability to maintain memories across sleep (e.g., Diekelmann & Born, 2010; Rasch & Born, 2013), is hardly tangible in age-comparative studies. If the definitions used to identify a given sleep stage do not capture the same phenomenon (e.g., the presence of slow oscillations) to the same degree across age groups, age differences in associations with outcome measures can hardly capture the intended mechanism. For example, the observation that memory performance is more consistently associated with SWS in younger than in older adults, could be taken as evidence that slow oscillations forfeit their functionality for memory processing during aging (e.g., Baran, Mantua, & Spencer, 2016; Scullin, 2012; Sonni & Spencer, 2015). However, if slow oscillations are present in older adults but just attributed to a different sleep stage, the statistics might be correct, but the functional interpretation would be different. In this case, slow oscillations found in a different sleep stage may continue to contribute to the consolidation of memories even in old age—a matter of fact only revealed when focusing on the oscillatory phenomena themselves (e.g., Helfrich, Mander, Jagust, Knight, & Walker, 2018; Mander et al., 2013; Muehlroth et al., 2019b; Varga et al., 2016; Westerberg et al., 2012).
Given the described challenges an NREM sleep subdivision bears, it may be hardly surprising that research in other species refrains from applying such a subclassification (but see Lacroix et al., 2018, for a recent attempt for “human-like” subclassification of NREM sleep in rodents). The term “SWS” is typically used to describe the whole continuum of NREM sleep, which has evoked pivotal misunderstandings when comparing human and animal research (Genzel et al., 2014). This conception of NREM sleep may arise from the prevailing sleep scoring procedure in rodents that can easily distinguish between REM and NREM sleep on the basis of movements and spectral parameters of electrical neuronal activity during sleep, but less so between NREM substages (Bagur et al., 2018; Lacroix et al., 2018). In humans, contemplating whole-night multitaper spectral EEG data, the distinct natures of REM and NREM sleep becomes evident irrespective of participants’ age, whereas differences between SWS and stage 2 sleep are more subtle as they transition gradually (Figure 2; see Prerau et al., 2017 for a detailed description of the spectral dynamics of the whole-night multitaper spectogram). Crucially, this does not signify that NREM sleep should be regarded as a uniform state, but rather that the continuous nature of NREM substages must be acknowledged, which makes a clear-cut distinction and definition of stage 2 and SWS problematic, if not impossible.

To conclude, an explicit distinction between stage 2 and SWS bears not only methodological but also theoretical difficulties. A coherent methodology to study the complex and dynamic nature of sleep not only within, but also between species is required to enable informative study comparisons (cf. Prerau et al., 2017). Research on the causes and consequences of interindividual and age-related differences in sleep physiology should thus transcend stage-based analyses and rather consider the actual exploranda, that is, the underlying neurophysiological processes (for instance, the occurrence of rhythmic thalamocortical sleep spindles) present during both stage 2 sleep and SWS.

4 | CHALLENGE 2: MULTIPLE WAYS TO DESCRIBE SLEEP PHYSIOLOGY

The multitude of indicators and available algorithms used to describe sleep-specific neural activity hampers the comparability of results between studies. We argue that the validity and value of age-comparative sleep research depends greatly on the precise definition of indicators that reflect a neural process of interest as directly as possible. In our view, indicators of sleep microstructure should always be preferred as they more closely mirror the dynamics and characteristics of sleep-specific neural activity. Everyone engaging in sleep research will soon realize the multitude and inconsistency of indicators available to describe neurophysiological activity during sleep (Figure S1). Comparing different measurement occasions (Brandmaier et al., 2018) and diverse indicators of slow oscillatory activity during NREM sleep in our own data set (namely % SWS, relative slow-wave activity [SWA], and the number, density, frequency, and amplitude of detected slow oscillations), we noticed a generally poor agreement between the different variables both across and within age groups ($ICC = 0.201 \, [-0.15; 0.47]$, $F(62,310) = 1.25, \, p = .113; \, ICC_{YA} = 0.13 \, [-0.48; 0.54]$, $F_{YA}(28,140) = 1.15, \, p_{YA} = .295; \, ICC_{OA} = 0.09 \, [-0.49; 0.49]$, $F_{OA}(33,165) = 1.1, \, p_{OA} = .342$; cf. Table S1). In contrast, the test–retest reliability of the same variable between two consecutively recorded nights was good to excellent (all $ICC \geq 0.78$, all $F \geq 4.61$, all $p < .001$; cf. Table S1). A moderate test–retest reliability of relative SWA in the group of older adults was the only exception ($ICC = 0.68 \, [.33; 0.85]$, $F(33,33) = 3.69, \, p < .001$). Critically, all participants engaged in an extensive associative learning task between the two recorded nights (cf. Muehlroth et al., 2019b, 2019a). As demanding learning is known to influence indicators of spindle and slow oscillatory activity (e.g., Gais, Mölle, Helms, & Born, 2002; Mölle, Bergmann, Marshall, & Born, 2011; Mölle, Eschenko, Gais, Sara, & Born, 2009), the reported estimates should even be considered as the lower bound of reliability. Taken together, even though sleep physiology can be very reliably measured by applying common analytic approaches, the great disparity of available indicators calls for a thorough consideration of the way we describe sleep physiology in our analyses.

In general, the validity of indicators to quantify neural activity during sleep scales with their resolution and precision: Age differences in sleep are best captured by describing the dynamics and features of neural activity in as much detail and as directly as possible. Using machine learning on 3,500 sleep recordings covering the ages 18–80, Sun and colleagues (2019) recently demonstrated that oscillatory dynamics during sleep (i.e., the microstructure of sleep)—like fluctuations in spectral power—can reliably predict participants’ age (mean absolute deviation 7.8 years; see Sun et al., 2019). Global sleep stage composition (i.e., the macrostructure of sleep), in contrast, turned out to be a less appropriate predictor of chronological age (mean absolute deviation 23.3 years; Sun et al., 2019). Sleep is known to be a dynamic process (Prerau et al., 2017; Spiess et al., 2018; Yetton, McDevitt, Cellini, Shelton, & Mednick, 2018). The use of sleep stage estimates mismatches our current theoretical understanding of sleep and strongly reduces resolution of the data (Conte & Ficca, 2013; Prerau et al., 2017). Altogether, we believe that the analysis of sleep EEG data should transcend beyond the
report of sleep stages and global spectral features and focus on the microstructure of sleep.

To avoid using the proportion of SWS as a proxy for the presence of slow waves, analyses increasingly report on slow-wave activity (SWA). SWA measures the spectral delta power (i.e., 0.5–4.5 Hz) via a transformation of the data in the frequency domain (e.g., a Fast Fourier Transformation, FFT). In conjunction with a normalization procedure—for example, dividing the respective frequency band by the total power of the whole spectrum—the influence of possible confounds such as frequency-unspecific interindividual differences in EEG amplitudes can be minimized (see, e.g., Dannhauer, Lanfer, Wolters, & Knösche, 2011; Frodl et al., 2001; Leissner et al., 1970; Segalowitz & Davies, 2004; Werkle-Bergner, Shing, Müller, Li, & Lindenberger, 2009, for arguments on how interindividual differences in anatomical properties of the brain and skull influence the EEG signal). A normalized measure of SWA increases the comparability across diverse study populations. However, spectral power, as conventionally measured with FFT, is a compound of rhythmic processes and the arrhythmic 1/f background spectrum that, crucially, is flattened in older compared to younger adults (Voytek et al., 2015; Figure S2; see Haller et al., 2018; Kosciessa et al., 2019; Wen & Liu, 2016; Whitten, Hughes, Dickson, & Caplan, 2011 for approaches to separate rhythmic activity from the arrhythmic background spectrum). Moreover, as spectral power conflates both duration and amplitude of rhythmic neural activity (e.g., Kosciessa et al., 2019), an interpretation of SWA in terms of abundance or strength of slow waves remains difficult (Amzica & Steriade, 1998; Mensen, Zhang, Qi, & Khatami, 2016). “Thus it is essential to strive toward a characterization of sleep EEG oscillations that faithfully represent the underlying data, allowing us to apply the wealth of knowledge we have gained about the continuum of the underlying neurophysiological mechanisms to the interpretation of the EEG” (Prerau et al., 2017, p. 63).

Technological developments over the past decades are increasingly facilitating analysis of specific features and characteristics of neural events during sleep (Mensen, Riedner, & Tononi, 2016; Mölle, Marshall, Gais, & Born, 2002; Prerau et al., 2017). Such analyses can incorporate the amplitude, oscillatory frequency, the density of occurrence (i.e., number of events per time interval), and the slope of specific oscillatory events. These different indicators are often used interchangeably although they signify differential physiological properties and, accordingly, show differential age dependency (cf. Box 2, Figure S1). If applied correctly, though, they hold great explanatory power and can provide more informed insights into the aging brain (Fogel et al., 2017; Sun et al., 2019; Ujma et al., 2019). Moreover, the possibility to identify individual oscillatory events also entails the opportunity to investigate how different oscillations and neural processes interact (Helfrich et al., 2018; Klinzing et al., 2016; Muehlroth et al., 2019b; Staresina et al., 2015). At the moment, due to the variety of algorithms available to study oscillatory brain signals and their interaction during sleep, these analyses are still time-consuming and between-study comparability of indicators can be challenging (Figure 3; cf. Warby et al., 2014). To tackle these problems, currently, several open source projects are under development that aim to harmonize and standardize analysis pipelines and parameters and make them publically available (e.g., Luna [http://zzz.bwh.harvard.edu/luna/], Sleep [Combrisson et al., 2017; http://visbrain.org/sleep.html], SPISOP [www.spisop.org], YASA [http://doi.org/10.5281/zenodo.3382284]; all code and sample data to reproduce this manuscript’s EEG analysis pipeline are available on the Open Science Framework [https://osf.io/zur2b/]). Identifying and resolving these discrepancies—ideally by the use of common algorithms—will lay the grounds for a mechanistic understanding of the causes and consequences of age-related changes in sleep (Helfrich et al., 2018; Muehlroth et al., 2019b).

5 CHALLENGE 3: AMPLITUDE REDUCTIONS—TOWARD ALGORITHMS WITH ADJUSTED THRESHOLDS

The benefits of a direct investigation of slow oscillatory events are constrained by a lacking consensus on the detection thresholds to be applied. We suggest that a fair assessment of true slow oscillatory activity requires the use of amplitude thresholds that are adjusted individually within participants.

Studying rhythmic neural activity during sleep requires a reliable detection of oscillatory events themselves. Here, automated slow oscillation detection methods offer an efficient way to identify the presence and characteristics of individual slow oscillations (e.g., Massimini, Huber, Ferrarelli, Hill, & Tononi, 2004; Mensen, Riedner, et al., 2016; Mölle et al., 2002; Nir et al., 2011). Typically, oscillation detection algorithms rely on a bandpass-filtered EEG signal within a theoretically implied frequency range (e.g., 0.5–1 Hz for slow oscillations, see Challenge 4 for details on the definition of frequency ranges). The presence of oscillatory components is then inferred by applying certain amplitude and peak detection criteria (cf. Supporting Information). Amplitude criteria can either be based on fixed thresholds or on variable thresholds adjusted within individuals (e.g., Massimini et al., 2004; Mölle et al., 2009, 2002; Staresina et al., 2015). In analogy to the visual scoring of SWS, many algorithms rely on a criterion of 75 μV (Carrier et al., 2011; Dubé et al., 2015; Piantoni et al., 2013) or even higher amplitudes (Heib et al., 2013; Massimini et al., 2004) for the detection of slow oscillations during
NREM sleep. But an increasing number of studies opt for thresholds adjusted individually for each participant (e.g., Klinzing et al., 2016; Mölle et al., 2011; Ngo et al., 2015). As yet, there is neither a gold standard nor consensus on the most suitable method nor on the threshold settings to be applied when identifying the presence of oscillatory events (Mensen, Zhang, et al., 2016; Ujma et al., 2019). However, the choice of detection thresholds has profound effects on the analysis of age differences.

Irrespective of the applied detection threshold (i.e., a fixed threshold of 75 μV vs. an individually adjusted threshold of 1.25 times the mean amplitude of all putative slow oscillations; cf. Supporting Information), in the older adults in our data, the number, density (i.e., the number of slow oscillations per minute of NREM sleep), amplitude, and frequency of slow oscillations were all reduced (Table 2). Nevertheless, compared to an individually adjusted amplitude criterion, the extent of age effects was augmented when using a fixed detection threshold of 75 μV ($d_{\text{adjusted}} = 0.67$ [0.14; 1.2], $d_{\text{fixed}} = 1.93$ [1.31; 2.56], nonoverlapping confidence intervals for Cohen's $d$ of slow oscillation density, cf. Table 2). Using individually adjusted thresholds, as displayed in Figure 1b, in older adults, the average slow oscillation amplitude fell below the 75-μV criterion in 61.77% of the subjects (21 out of 34 compared to only one younger subject [3.45%]). Using individually adjusted versus fixed detection thresholds, as displayed in Figure 1b, in older adults, the average slow oscillation amplitude fell below the 75-μV criterion in 61.77% of the subjects (21 out of 34 compared to only one younger subject [3.45%]). In other words, in older adults, a large number of potential slow oscillations will not be detected in case of a fixed-amplitude criterion of 75 μV as the detection is biased toward higher-amplitude oscillations (Figure 3b). As a consequence, slow oscillation density estimates that are derived using individually adjusted versus fixed detection thresholds are comparable in younger but not in older adults ($Z_{FA} = -0.88$, $p_{FA} = .38$; $Z_{OA} = 6.08$, $p_{OA} < .001$; see Table 2; Figure 3a).

Critically, interindividual differences in slow oscillation amplitudes might not only carry functional information but may be confounded by differences in homeostatic sleep.
regulation (Dijk et al., 2000), brain integrity (Dubé et al., 2015; Saletin et al., 2013), and skull thickness (Frodl et al., 2001; Leissner et al., 1970). Hence, an individually adjusted amplitude threshold that tags slow oscillations with the largest amplitudes within individuals may have advantages. Only when slow oscillations are selected independently of inter-individual differences in EEG amplitudes, do unbiased comparisons become possible. Persisting age differences in the amplitude, density, and morphology of slow oscillations can then be interpreted as valid age group differences. Only when slow oscillations are selected independently of inter-individual differences in EEG amplitudes, do unbiased comparisons become possible. Persisting age differences in the amplitude, density, and morphology of slow oscillations can then be interpreted as valid age group differences. Compared to fixed-amplitude detection criteria, especially in age-comparative settings, the use of individually adjusted amplitude criteria for detecting slow oscillations during NREM sleep provides a more realistic and at the same time less biased measure of true slow oscillatory activity.

**FIGURE 3** Comparison of fixed (75 µV) and individually adjusted slow oscillation detection thresholds. (a) Relation between slow oscillation density estimates derived by the use of individually adjusted versus fixed detection thresholds. In younger adults (orange), the two detection thresholds produce comparable results. In older adults (blue), especially for individuals with small slow oscillation amplitudes (smaller square size), a fixed detection threshold results in an underestimation of slow oscillation density. (b) Individual average slow oscillations and age group averages of slow oscillations detected with individually adjusted and fixed amplitude thresholds. Left panel: In younger adults, both algorithms result in comparable slow oscillation detection. Right panel: In older adults, individually adjusted thresholds (blue) result in the detection of slow oscillations with smaller amplitudes compared to higher-amplitude fixed-threshold detection (green). OA, older adults; SO, slow oscillation; YA, younger adults

### 6 | CHALLENGE 4: DIFFERENTIAL FREQUENCY SHIFTS IN SLEEP OSCILLATIONS—TOWARD INDIVIDUALIZED FREQUENCY RANGES

The exact frequency band in which an oscillatory event can be observed varies between individuals and is influenced by age. To prevent missing true oscillatory events during detection and mixing up functionally distinct oscillatory components, cutoff frequencies have to be determined carefully. We argue that individually determined frequency ranges are the prerequisite to derive meaningful age- and person-specific characteristics of oscillatory components.
6.1 Contemplating frequency cutoffs for low-frequency oscillatory components

Age-related changes in the appearance of sleep oscillations do not only include alterations in amplitudes, but also entail shifts in the intrinsically expressed frequencies of the respective neural events. With respect to low-frequency components, a global slowing has consistently been reported (Carrier et al., 2011; Dubé et al., 2015; Mander et al., 2017) (Figure 4c, Box 2). These effects comprise both slow oscillations (<1 Hz) as well as slow waves with higher frequencies (1–4 Hz). Although slow oscillations and slow waves constitute different entities on a theoretical level (Achermann & Borbély, 1997; Amzica & Steriade, 1998; Steriade, Nunez, & Amzica, 1993), the applied cutoff frequencies defining low-frequency EEG components are often not much attended to.

Empirically, the problem of identifying the exact frequency ranges for specific oscillatory components in individual participants can be solved by considering the NREM sleep power spectrum. Peaks in the power spectrum signify the presence of rhythmic neural activity at a given frequency (Aru et al., 2015; Kosciessa et al., 2019; Whitten et al., 2011). When visually inspecting the age-specific power spectra during NREM sleep in our data, it became evident that the peak in low-frequency power was shifted to even lower frequencies in older adults (Figure S2). This slowing in low frequencies was statistically reliable (average frequency peak: t(115) = 6.18, p < .001, MYA = 0.73 Hz, MOA = 0.53 Hz) and was also evident for discrete slow oscillation events (Z = 3.24, p = .001, MdYA = 0.79 Hz, MdOA = 0.77 Hz; see Table 2, Figures 3b, 4c). Although statistically detectable, in contrast to the dramatic reduction in slow oscillation amplitudes, interindividual and age differences in slow oscillation frequencies were relatively small (d_frequency = 0.90 [0.36; 1.44], d_amplitude = 2.28 [1.62; 2.94], nonoverlapping confidence interval; cf. Table 2).

Taken together, it appears that the magnitude of the slowing of slow oscillations in old age does not reach an extent that would bias detection results when established frequency ranges (e.g., 0.5–1 Hz for slow oscillations) are utilized. Nonetheless, we emphasize the need to carefully consider, justify, and report the selected frequency range when publishing sleep research findings on older adults. Slow oscillations and “faster” slow waves have been associated with differential homeostatic regulation (Bersaglieri & Achermann, 2010; Werth, Achermann, Dijk, & Borbe, 1997) as well as with divergent roles in the context of cognitive pathology in aging (Lucey et al., 2019; Mander et al., 2015, 2016). Yet, detection algorithms often deploy frequency bands that comprise both oscillatory components (e.g., Carrier et al., 2011; Kurth et al., 2010). Finally, one should bear in mind that even when using adequate frequency cutoffs, filtering can induce signal distortions. The choice of filter type and filter parameters will differentially influence the signal and change detection results (Widmann, Schröger, & Maess, 2015). Together, we thus underscore the need for clear and coherent terminology along with appropriate parameter settings and detection criteria when reporting on age differences in low-frequency oscillatory activity.

6.2 Interindividual (age) variation in fast and slow spindle frequency bands

In contrast to slow oscillations whose detection seems mostly dependent on appropriate amplitude settings, the detection of sleep spindles is heavily influenced by the selection of frequency ranges. In particular, the detection of spindle events is challenged by the presence of two spindle types. Slow (ca.

![Figure 4](image-url)  
**Figure 4** Topographical distribution of the average slow oscillation and spindle frequencies. (a) Overall increased fast spindle frequency in older (blue) compared to younger (orange) adults, with reduced effects at frontal channels. (b) Decreases in slow spindle frequency in older adults are only pronounced at Cz. (c) Uniform decrease in slow oscillation frequency in old age at all channels. Note. All slow oscillation and spindle characteristics were derived using individually adjusted amplitude thresholds. OA, older adults; SO, slow oscillation; YA, younger adults.
9–12.5 Hz) and fast spindles (ca. 12.5–16 Hz) differ in frequency, topography (De Gennaro & Ferrara, 2003; Doran, 2003; Klinzing et al., 2016; Schabus et al., 2007), generating mechanisms (Ayoub et al., 2013; Schabus et al., 2007; Timofeev & Chauvette, 2013), and not least in their functionality (Barakat et al., 2011; Fogel & Smith, 2011; Mölle et al., 2011). Although the existence of two spindle types is now widely acknowledged, there is no general consensus on the exact definition of the frequency range in which to expect slow and fast spindles (Cox, Schapiro, Manoach, & Stickgold, 2017). Critically, spindles undergo a complex developmental shift and differentiation in their inherent frequency, power, and topography (Campbell & Feinberg, 2016; Crowley et al., 2002; Nicolas et al., 2001; Purcell et al., 2017). This not only challenges research on younger adults (Cox et al., 2017), but also research that focuses on the entire lifespan.

In line with these observations, slow and fast spindle frequencies were differentially affected by age in our data. Average fast spindle frequency was increased in older adults ($F(1,56) = 14.12, p < .001, M_{YA} = 13.39$ Hz, $M_{OA} = 13.68$ Hz) (Figure 4a, Box 2). In contrast, an age-related decrease in the average frequency of slow spindles was only pronounced at electrode Cz (age by channel interaction: $F(3,168) = 9.78, p < .001, Z_{Cz} = 3.31, p < .001, M_{d_{YA}} = 11.28$ Hz, $M_{d_{OA}} = 10.61$ Hz; all other $ZI \leq 1.41$, all other $p \geq .158, \alpha$-level adjusted to .0125; Figure 4b). The complexity of frequency shifts during aging is thus even enhanced by the topographical intricacy of the observed effects.

The multiple facets of age differences in sleep spindle appearance require the careful consideration of an age-adjusted definition of spindle frequency ranges. Based on individually identified peaks in the NREM sleep power spectrum at topographical locations where slow and fast spindles are expected (cf. Figure S2), individually adjusted spindle frequency ranges can be defined—even if they might not match conventionally prescribed bands exactly (Adamczyk, Genzel, Dresler, Steiger, & Friess, 2015; Bódizs, Körmendi, Rigó, & Lázár, 2009; Cox et al., 2017; Doppelmayr, Klimesch, Pachinger, & Ripper, 1998; Grandy et al., 2013; Klimesch, 1999; Ujma et al., 2015). Comparisons between spindle detection algorithms that use commonly predefined frequency ranges and those that adjust their frequency range definitions within individuals, provide a strikingly poor agreement with conventional (fixed-frequency) algorithms. The latter have been shown to miss up to 25% of fast and slow spindles (Adamczyk et al., 2015; Cox et al., 2017; Ujma et al., 2015).

To allow for the investigation of differences in spindle characteristics, age-comparative research should ensure that the frequency bands applied to define slow and fast spindles are congruent with age- and person-specific spindle properties. Moreover, frequency definitions need to be broad enough to capture sleep spindles in their complexity and entirety. Independently of an individual’s age, person-specific definitions of spindle frequencies form the basis for these improvements (Adamczyk et al., 2015; Cox et al., 2017; Ujma et al., 2015).

7 | CHALLENGE 5: TOPOGRAPHICAL DISPARITIES IN AGE-RELATED SLEEP CHANGES

Age-related alterations in sleep physiology do not show a spatially coherent pattern. We highlight that alterations in sleep oscillations that are measurable on the scalp do not necessarily reflect an altered initiation of underlying physiological sleep processes per se, but could rather represent age-related alterations in structural properties of the brain. Taking into account that intrinsic characteristics of sleep-specific oscillatory components may be well preserved during aging can help our understanding of the functionality of sleep in old age.

Across the lifespan, the human brain undergoes complex structural alterations (Buckner, 2004; Cabeza, Nyberg, & Park, 2005; Fjell & Walhovd, 2010; Fjell et al., 2013; Raz et al., 2005), which are also reflected in the topography of the sleep EEG (Buchmann et al., 2011; Carrier et al., 2011; Kurth et al., 2010; Landolt & Borbély, 2001; Martin et al., 2013; Sprecher et al., 2016). From childhood to adolescence, along with cortical maturation in frontal brain regions, maximal SWA shifts from posterior to frontal brain regions (Kurth et al., 2010; Whitford et al., 2007). In old age, frontal brain regions are particularly prone to age-related structural white and gray matter loss (Giorgio et al., 2010; Raz & Rodrigue, 2006). As a consequence, age-related decline in spectral EEG power and amplitude is especially pronounced over frontal areas (cf. Figure S7, Table S6) (Carrier et al., 2011; Landolt & Borbély, 2001; Mander et al., 2017; Martin et al., 2013; Sprecher et al., 2016).

Topographical shifts in sleep parameters are diverse. For example, in our own data, the centroparietal prevalence of fast spindle density and amplitude in younger adults was shifted posterior in older adults, resulting in reduced or even absent age differences over parietal and occipital regions (Figure 5a,d; Figure S2). In contrast, slow spindles and slow oscillations maintained their frontal dominance in old age. Whereas the amplitude and density of slow oscillations were globally and uniformly reduced in old age (Figure 5c,f), the density and amplitude of slow spindles showed most pronounced age-related reductions over frontal areas (Figure 5b,e). Thus, the extracted oscillatory event or indicator (i.e., slow spindle, fast spindle, and slow oscillation), their derived characteristics (i.e., density, amplitude, and frequency), and not least an individual’s age interact and influence where, and to what extent, age differences in sleep can be observed.
Crucially, the topographical heterogeneity of age effects detectable on the scalp level may be meaningful and indicative of alterations in underlying brain structure and cellular properties. To illustrate this effect, we tested for the association between gray matter volume and fast spindle density measured at different electrode sites (separate voxel-wide multiple regression models for each predictor; see Figure 6; cf. Supporting Information; Figure S4, Table S3). In line with the observed posterior preservation of fast spindles in older adults (Figure 5a,d; Figure S2), fast spindle density at Pz and Oz did not relate to gray matter volume. In contrast, fast spindle density measured frontally and at Cz, that was most reduced on older adults, was associated with structural integrity in frontal and temporal brain regions (Figure 6; Figure S4, Table S3). In comparison, gray matter clusters for the association with slow spindle and slow oscillation density were small and less consistent (Figures S5 and S6, Tables S4 and S5). Based on these analyses examples, it becomes clear that structural brain integrity does not globally shape the generation and propagation of all oscillatory components. However, one should note that chronological age, in general, was associated with global reductions in gray matter volume that were strongest in frontal brain regions (Figure 6; Figure S3, Table S2). The extent of structural brain alterations in aging can thus overshadow more focalized and potentially meaningful structure–function associations.

To derive a meaningful interpretation of the observed structure–function associations, one needs to consider that the extent to which we can detect oscillatory events on the scalp depends on both their generation and propagation. Thus, we can encounter the situation that oscillations are properly generated, but their propagation to certain brain areas is impaired by altered structural properties of the involved brain circuits (Landolt & Borbély, 2001; Martin et al., 2013; Nir et al., 2011; Sprecher et al., 2016). In line with this notion and the gray matter associations reported above, the maximum of fast spindle density over the posterior cortex has been interpreted in terms of aggravated fast spindle propagation to frontal and central brain regions that is severely affected by brain atrophy during aging. In contrast, the generation of sleep spindles is often regarded as less affected during aging (Crowley et al., 2002; Landolt & Borbély, 2001; Martin et al., 2013; Nir et al., 2011; Sprecher et al., 2016).
2011; Sprecher et al., 2016; see also Box 2). Still, we must acknowledge that the available cross-sectional correlative data neither allows us to disentangle whether differences in oscillatory components stem from disparities in their generation or their propagation, nor whether structural brain alterations cause, accompany, or result from alterations in sleep physiology. To solve these questions, longitudinal study designs that identify and track the sources of oscillatory components and link them to aging changes in structural brain integrity are required.

To conclude, the topographic distinctiveness of age-related sleep alterations is an important but often overlooked factor in age-comparative sleep studies. With the choice where to study age differences, age effects can either be maximized or reduced. Hence, deriving measures of EEG activity from only one derivation, as common in standard sleep recordings, should be avoided (Mander et al., 2017). To capture regional variation of age effects in the EEG, electrode setups that can capture hemispherical asymmetry (Sprecher et al., 2016) and allow for a high spatial resolution should be preferred (ideally at least a 32-electrode setup or even high-density EEG). The fact that sleep parameters are selectively altered over specific brain regions, but not over other areas, casts doubt on the reported extent of age-related impairments in sleep processes. Brain atrophy likely results in a diminished and altered generation of certain sleep parameters (Dubé et al., 2015; Mander et al., 2013), but may also affect the propagation of electrical events that, per se, may be less prone to age-related changes. Future research should interpret the reported effects with care and provide solutions to disentangle age-related impairments in slow wave and spindle generation and propagation. Rather than placing a continued focus on the deficiency of sleep in old age, sleep studies require a shift toward maintained and intact sleep characteristics to determine altered or preserved functionality of sleep in old age.

**FIGURE 6** Age differences in fast spindle density mirror cortical integrity. (a) Cohen’s d of age differences in fast spindle density at different electrode sites. Confidence intervals are displayed as error bars. Age differences are most pronounced at Cz but are attenuated over more posterior sites. (b) Sagittal, coronal, and axial depiction of the association between gray matter volume and age (yellow), frontal (violet), and Cz (cyan) fast spindle density (separate voxel-wise multiple regression models for each predictor, controlled for total intracranial volume). In contrast to a global age-related decrease in gray matter volume (see Figure S3, Table S2), associations with spindle density are more specific. Whereas interindividual differences in fast spindle density at Pz and Oz are not related to gray matter volume, fast spindle density at Cz shows a strong association with structural integrity in frontal and temporal brain regions (see Figure S4, Table S3). Clusters are displayed and considered significant at a FWE-corrected voxel threshold of $p < .05$, and a FWE-corrected cluster extent threshold of $p < .05$ and $k > 100$. All clusters are corrected for nonstationary smoothness. See Tables S3–S5 for the cluster statistics and Figures S4–S6 for detailed depictions of the clusters. FWE, family-wise error; OA, older adults; YA, younger adults.
Box 3 Methodological advice for age-comparative sleep studies

1. Closely inspect your sleep EEG data visually. Knowing what the data look like will ease your understanding of what to consider and what to adjust in your analyses.
2. Analyses should focus on microstructural rather than macrostructural sleep features.
3. Define the neural process of interest as well as possible. Avoid proxies wherever possible.
4. Consider that seemingly similar measures (e.g., frequency, amplitude, density, power of the very same spindle event) can reflect very different physiological and neuronal properties. Be precise on which features you extract and interpret them carefully.
5. NREM sleep forms a continuum. Acknowledge that neurophysiological processes that are captured by sleep spindles and slow oscillations are present during both stage 2 sleep and SWS.
6. Use individually adjusted algorithms (with regard to amplitude and frequency) for sleep oscillation detection.
7. Use electrode setups that enable topographical comparisons. Try to incorporate the mechanisms that are driving topographical heterogeneity in age effects into your theories and methods (e.g., by carrying out structural brain imaging and/or source localization techniques in addition).
8. Instead of focusing on deficits only, consider what is intact or maintained in old age. Do your very best to apply age-adjusted methodology and comparisons.

8 | CONCLUSION

Fundamental changes in human sleep during aging are consistently reported and easy to observe—but, as discussed in this article, it is nontrivial to quantify and measure them. Intact sleep and the functional succession of different sleep states is enabled by the balance of different neurotransmitter systems (Aston-Jones, Gonzalez, & Doran, 2007; Cirelli, Huber, Gopalakrishnan, Southard, & Tononi, 2005; Ramm, 1979; Steriade, Datta, Paré, Oakson, & Curró Dossi, 1990), and synergistic activity in different cortical and subcortical brain regions (Hobson & Pace-Schott, 2002; Pace-Schott & Hobson, 2002; Saletin et al., 2013; Steriade et al., 1990). Furthermore, it is embedded in a system of regulatory circadian and homeostatic processes (Achermann & Borbély, 1999; Dijk et al., 2000). Altogether, the diversity of regulatory mechanisms, reaching from genes and cells to neurotransmitters and brain networks (Hobson & Pace-Schott; Pace-Schott & Hobson, 2002), makes sleep susceptible to a wide range of possible interference. In consequence, the causes and the appearance of age-related changes in sleep are manifold and diverse (Dubé et al., 2015; Fogel et al., 2012, 2017; Mander et al., 2017; Muehlroth et al., 2019b; Zhong et al., 2019).

Standard criteria applied to measure these alterations are commonly optimized to capture interindividual variance in healthy younger adults but can be severely biased when applied in other populations. As a result, not all age differences documented via the use of conventional analysis pipelines capture real alterations in underlying neural processing. We thus stress the need to develop individually adjusted analytic approaches in order to disentangle true age-dependent effects from age-independent interindividual differences (see Box 3). But how could such an age-adjusted research process work?

We suggest that the basis for research on age-dependent differences in a given (neural) process is rooted in thorough theoretical considerations that incorporate the complex and dynamic nature of aging (e.g., Cabeza et al., 2018; Cabeza, Nyberg, & Park, 2005; Lindenberger, Li, & Bäckman, 2006) and the individual as the primary unit of analysis (Fisher, Medaglia, & Jeronimus, 2018; Grandy, Lindenberger, & Werkle-Bergner, 2017; Molenaar & Campbell, 2009; Nesselroade, Gerstorf, Hardy, & Ram, 2007; Nesselroade & Molenaar, 2016; Rose, Rouhani, & Fischer, 2013). Ideally, such a research process starts out with an exact definition of a neural or physiological phenomenon that we intend to investigate. For instances, we could ask whether the intensity of sleep (in terms of sleep depth) is reduced in old age. In a next step, potential indicators that mirror the phenomenon of interest along with its crucial features have to be considered and the most appropriate indicator(s) need(s) to be selected. In our example, we could decide to use SWA and slow oscillation density as markers of low-frequency oscillatory activity during sleep that index the depth of sleep (e.g., Achermann & Borbély, 1999; Bersagliere & Achermann, 2010; Blake & Gerard, 1937). Next, we need to determine how to best measure the respective indicator. Here, the individual has to be at the center of the analysis. Interindividual differences that are irrelevant for our research question and thus introduce “noise” (e.g., differences in anatomical skull properties) should be filtered out by adjusting analytic detection criteria at the level of the individual person (Karch, Sander, von Oertzen, Brandmaier, & Werkle-Bergner, 2015; Molenaar & Campbell, 2009; Nesselroade et al., 2007; Nesselroade & Molenaar, 2016; Rose et al., 2013). In this way, we can create the conditions that facilitate an accurate interpretation of remaining interindividual differences (Nesselroade et al., 2007). To be as unbiased as possible when comparing heterogeneous subject groups and individuals, we thus
believe that research on sleep and aging should strive to identify the best possible analysis strategy to study a certain phenomenon of interest—even if this means moving away from conventional analytic approaches.

Complementarily, we stress the inevitable need to shift research foci from cross-sectional toward longitudinal studies. The currently available evidence cannot inform about “real” age changes in sleep (Hertzog & Nesselroade, 2003; Hofer & Sliwinski, 2001; Li & Schmiedek, 2002; Lindenberger et al., 2011; Lindenberger & Pöpper, 1998; Overton, 2010; Raz & Lindenberger, 2011). To disentangle adaptive from maladaptive aging processes, and to identify the lead-lag relations between age-related alterations in brain structure and sleep, study designs that follow the same individual across multiple measurement occasions are required (Hofer & Sliwinski, 2001; Li & Schmiedek, 2002; Lindenberger, 2014; Raz & Lindenberger, 2011).

Aside from that, we want to emphasize that age-dependent alterations in sleep physiology should not necessarily be interpreted in terms of “shortage” or “deficiency.” The brain is an adaptive organ that changes its structure and function in reaction to mismatches between environmental demands and organismic supplies (Lövdén, Backman, Lindenberger, Schaefer, & Schmiedek, 2010; Park & Reuter-Lorenz, 2009). First, this means that not all observed age differences in sleep necessarily indicate a loss (Baltes, 1987; Baltes, Staudinger, & Lindenberger, 1998). Reduced SWA, for instance, might merely be the consequence of reduced sleep pressure and an altered need for sleep in old age (Dijk et al., 2000). Second, due to the adaptive nature of the human brain, persons with very different developmental trajectories and physiological conditions can display similar sleep patterns (Muehlroth et al., 2019a). By identifying the similarities among individuals, and specifically among younger and older adults (Nesselroade et al., 2007), we can determine whether and how sleep is preserved in old age. Moreover, we can determine whether maintained sleep physiology is also reflected on the level of behavioral outcomes (Cabeza et al., 2018; Muehlroth et al., 2019b, 2019a; Nyberg, Lövdén, Riklund, Lindenberger, & Bäckman, 2012; Nyberg & Pudas, 2019; Park & Reuter-Lorenz, 2009). Research that shifts its focus toward the successfull retention of sleep characteristics represents the foundation to gain a mechanistic understanding of the consequences entailed in altered sleep physiology. This is the basis for identifying the diagnostic value of sleep changes (Ferrarelli et al., 2019; Grandy et al., 2013; Prinz, 1977).

Finally, we want to point out that understanding sleep neurophysiological dynamics in their entirety ultimately requires taking a holistic view on sleep as alternating states of NREM and REM sleep (Conte & Ficca, 2013; Prerau et al., 2017; Scullin & Gao, 2018). Beyond a continued focus on NREM sleep, theories and analysis methods should take the cyclic nature of NREM and REM sleep into account and consider their dynamic interaction (e.g., by investigating stage transition rates, probabilities, and their temporal pattern; cf. Kishi et al., 2011; Schlemmer, Parlitz, Luther, Wessel, & Penzel, 2015; Yetton et al., 2018).

Recent attempts to establish sleep as a novel biomarker and treatment target for Alzheimer’s disease have instantly evoked a massive rise in research on sleep and aging (Ju et al., 2014; Mander et al., 2016; Noble & Spires-Jones, 2019; Vaou et al., 2018). Methodological considerations, though, call for a clarification of what we want to measure and how we can best assess the respective phenomena. Improved methodological approaches will clearly speed up the search for a mechanistic understanding of the association between sleep and cognition in aged individuals.

ACKNOWLEDGMENTS

This work was financed by the Max Planck Society and grants from the German Research Foundation (DFG, WE 4269/3-1, WE 4269/5-1) to MWB. MWB’s work was supported by the Jacobs Foundation (Early Career Research Fellowships). BEM was supported by the Max Planck International Research Network on Aging, and the Max Planck School of Cognition. The authors thank Ann-Kathrin Joechner for helpful comments on an earlier version of the manuscript and Julia Delius for editorial assistance. We are grateful to everyone who contributed to the collection of data utilized in this article.

ORCID

Beate E. Muehlroth https://orcid.org/0000-0002-1064-4841
Markus Werkle-Bergner https://orcid.org/0000-0002-6399-9996

REFERENCES


Beate E. Muehlroth https://orcid.org/0000-0002-1064-4841
Markus Werkle-Bergner https://orcid.org/0000-0002-6399-9996

REFERENCES


Beate E. Muehlroth https://orcid.org/0000-0002-1064-4841
Markus Werkle-Bergner https://orcid.org/0000-0002-6399-9996

REFERENCES


Mensen, A., Zhang, Z., Qi, M., & Khatami, R. (2016). The occurrence of individual slow waves in sleep is predicted by heart rate. *Scientific Reports, 6*, 29671. https://doi.org/10.1038/srep29671


to sleep. *NeuroImage*, 178, 23–35. https://doi.org/10.1016/j.neuroimage.2018.05.024


**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Muehlroth BE, Werkle-Bergner M. Understanding the interplay of sleep and aging: Methodological challenges. *Psychophysiology*. 2020;57:e13523. https://doi.org/10.1111/psyp.13523