Received: 16 September 2023 | Revised: 22 January 2024 | Accepted: 23 January 2024

DOI: 10.1111/jsr.14163

REVIEW ARTICLE

Guidelines for the development, performance evaluation and validation of new sleep technologies (DEVSleepTech guidelines) – a protocol for a Delphi consensus study

Gabriel Natan Pires[1,2,3] | Erna S. Arnardóttir[4,5] | Sébastien Bailly[6] | Walter T. McNicholas[7,8]

[1]Departamento de Psicobiologia, Universidade Federal de São Paulo, São Paulo, Brazil

[2]Sleep Institute, São Paulo, Brazil

[3]European Sleep Research Society (ESRS), Regensburg, Germany

[4]Reykjavik University Sleep Institute, Reykjavik University, Reykjavik, Iceland

[5]Landspitali, The National University Hospital of Iceland, Reykjavik, Iceland

[6]Grenoble Alpes University, Inserm U1300, Grenoble Alpes University Hospital, Grenoble, France

[7]School of Medicine and the Conway Research Institute, University College Dublin, Dublin, Ireland

[8]Department of Respiratory and Sleep Medicine, St Vincent's Hospital Group, Dublin, Ireland

Correspondence
Gabriel Natan Pires, Departamento de Psicobiologia, Universidade Federal de São Paulo, Rua Napoleão de Barros, 925, CEP: 04024-002, São Paulo, Brazil.
Email: gabriel.pires@unifesp.br

Funding information
European Union's Horizon 2020 research and innovation program, Grant/Award Number: 965417

Summary

New sleep technologies are being developed, refined and delivered at a fast pace. However, there are serious concerns about the validation and accuracy of new sleep-related technologies being made available, as many of them, especially consumer-sleep technologies, have not been tested in comparison with gold-standard methods or have been approved by health regulatory agencies. The importance of proper validation and performance evaluation of new sleep technologies has already been discussed in previous studies and some recommendations have already been published, but most of them do not employ standardized methodology and are not able to cover all aspects of new sleep technologies. The current protocol describes the methods of a Delphi consensus study to create guidelines for the development, performance evaluation and validation of new sleep devices and technologies. The resulting recommendations are not intended to be used as a quality assessment tool to evaluate individual articles, but rather to evaluate the overall procedures, studies and experiments performed to develop, evaluate performance and validate new technologies. We hope these guidelines can be helpful for researchers who work with new sleep technologies on the appraisal of their reliability and validation, for companies who are working on the development and refinement of new sleep technologies, and by regulatory agencies to evaluate new technologies that are looking for registration, approval or inclusion on health systems.

KEYWORDS
consensus, Delphi, digital health, digital medicine, guidelines, mobile applications, sleep technology, sleep trackers, smartphones, smartwatches, wearables

1 | INTRODUCTION

Sleep medicine has always been an innovative field, and since its early years it has taken advantage of technological developments, which allowed a better understanding of the physiology of sleep and the pathophysiology of its disorders (Penzel et al., 2021). New sleep technologies are now being developed, refined and made available at a pace never seen before (Perez-Pozuelo et al., 2020; Pires et al., 2023; Schutte-Rodin et al., 2021). Wearable and nearable devices, bed sensors and sleep-related smartphone applications (apps) are becoming increasingly popular, being used for different purposes, including screening, diagnosis, treatment and follow-up of sleep-related conditions. Their intended user group also vary, with devices being directed to healthcare professionals, to the consumers directly (i.e. consumer-sleep technologies [CST]) or as hybrid technologies (which may have functionalities to both groups; Schutte-Rodin et al., 2021).

The surge on the number of new sleep technologies might be explained by three main factors: epidemiological, practical and commercial aspects (Pires et al., 2023). From an epidemiological perspective, the prevalence of sleep complaints and sleep disorders is high and increasing (Benjafield et al., 2019; Ferrie et al., 2011; Santos-Silva et al., 2010). Such high prevalence is responsible for creating the demand for screening, diagnosis and treatment of sleep disorders. The practical aspects are related to the current limitations on the diagnosis and treatment of sleep disorders. Diagnostics and treatments in sleep medicine are usually subjected to high costs, low availability and poor patient experience (Pires et al., 2023). These limitations reinforce the need to invest in more affordable, accessible and user-friendly diagnostics and therapeutics. Finally, the commercial aspects are explained by the combination of high demand and the limitations of current diagnostic and therapeutic methods. These circumstances have been an important drive for companies (including startups) to invest, innovate, design, develop and improve sleep technologies. Therefore, the market of sleep-related products is growing by ~18% per year, and reached ~USD2 billion in 2019 (Nester, 2019).

For the reasons above, especially regarding the practical limitations for currently available diagnostics and therapeutics, new sleep technologies are needed, providing that they are reliable, accurate, safe, and have been properly tested and validated. However, there are serious concerns about the validation and accuracy of new sleep-related technologies being made available (de Zambotti et al., 2020). Most of them, especially CSTs, have not been tested in comparison with gold-standard methods or have been approved by health regulatory agencies (Behar et al., 2013; Fino & Mazzetti, 2019; Khosla et al., 2018). As an example, only ~3% of all apps relating to obstructive sleep apnea seem to have been properly validated (Baptista et al., 2022). Another important limitation is the lack of transparency on the process of development or validation of new sleep technologies, which are often not properly disclosed or published (Goldstein & Depner, 2021; Schutte-Rodin et al., 2021). The lack of standards on the development, validation, performance evaluation and registration of new sleep technologies impairs a proper evaluation of their accuracy and reliability, thus questioning their actual usefulness (Baptista et al., 2022; Fino & Mazzetti, 2019).

The importance of proper validation and performance evaluation of new sleep technologies has already been discussed in previous studies (Baptista et al., 2022; de Zambotti et al., 2022; Depner et al., 2020; Fino & Mazzetti, 2019; Khosla et al., 2019), and some recommendations have already been published (de Zambotti et al., 2022; Khosla et al., 2019; Menghini et al., 2021; Schutte-Rodin et al., 2021; Tangudu et al., 2021). Although useful as initial steps towards the standardization of the process of development, validation and performance evaluation of new sleep technologies, these studies lack a systematic approach, do not employ standard methodology for the development of guidelines, and are not able to cover all aspects of new sleep technologies.

The current protocol describes the methods of a modified Delphi consensus study to elaborate guidelines for the development, performance evaluation and validation of new sleep devices and technologies (hereinafter, DEVSleepTech guidelines). It will be based on the guidance from EQUATOR (Enhancing the QUAlity and Transparency Of health Research) for developers of health research reporting guidelines (Moher et al., 2010) and on the CREDES (Guidance on Conducting and REporting DElphi Studies) guidelines (Jünger et al., 2017).

2 | DEFINITION OF NEW SLEEP TECHNOLOGIES

Although there is no clear, unanimous, unambiguous or official definition of what constitutes "new sleep technologies", the concept is reasonably well understood. It has been evaluated by studies and efforts from different societies and institutions as a way to bring together all new devices, apps and other forms of technologies that have been developed and refined in sleep medicine. It includes the #SleepTechnology resource from the American Academy of Sleep Medicine (AASM), and the SleepTech Portal and Award by the National Sleep Foundation. The definition of "new sleep technologies" used for the purposes of this work is based on and adapted from previous definitions and descriptions (Khosla et al., 2018; Khosla et al., 2019; Schutte-Rodin et al., 2021), being subjected to the criteria disclosed below.

2.1 | Current applicability

- *Inclusion*: Technologies that are currently neither considered as gold-standard methods nor are officially recommended by major sleep-related societies as a primary choice for the monitoring and scoring of sleep, or for diagnostic, screening, follow-up and treatment of sleep disorders and other sleep-related conditions.
- *Exclusion*: Gold-standard methods widely implemented for diagnostic, screening, follow-up and treatment of sleep disorders and other sleep-related conditions.

2.2 | Level of innovation

- *Inclusion*: Ranging from important improvements and modifications of standard technologies to highly innovative and disruptive sleep-related technologies and solutions.
- *Exclusion*: Minor modifications of standard technologies that require no further validation or that does not significantly alter its usage or implementation.

2.3 | Intended public

- *Inclusion*: It includes technologies that are intended to be used by or under supervision of healthcare professionals (clinical-grade technologies), or directly by the consumer (CST). It also includes

intermediate solutions that lie between these two groups (i.e. hybrid and transitional technologies; Schutte-Rodin et al., 2021).

- *Exclusion*: No exclusion criteria related to the intended public.

2.4 | Format

- *Inclusion*: Any technology that is mainly based on hardware or software. It includes devices for the monitoring and scoring of sleep, the improvement of sleep in a wellness perspective, and for diagnosis, screening, follow-up and treatment of sleep disorders in different ways, including wearables, nearables and bed sensors, smartphone apps, software and algorithms. It also includes innovation on data analysis, including the implementation of techniques related to artificial intelligence (including machine learning and deep learning) and other techniques for evaluation of sleep-related information.
- *Exclusion*: Development of sleep technologies that are not primarily related to a hardware or a software are excluded. It includes development of pharmacological interventions, surgical interventions, questionnaires and other subjective sleep assessment tools, among others. Exceptions are made to cases in which any of these items are implemented by innovative means or entails new technologies (e.g. sleep questionnaires integrated into an app or analysed through machine learning, an online sleep log that provides personalized feedback based on artificial intelligence, surgical interventions that apply innovative devices).

The evaluation, registration or compliance to regulatory requirements and agencies is an important aspect from the development to commercialization of new sleep technologies (e.g. Medical Devices Regulation [MDR] in Europe, or related regulations by the American National Standards Institute [ANSI] and the Food and Drug Administration [FDA] in the USA). However, they are not considered as part of the definition of new sleep technologies under the scope of this study, for two reasons. First, these guidelines are intended to assist in the development of new sleep technologies from the earliest steps, including the conceptualization and minimal viable product phases. Technologies in these phases are usually experimental, preliminary and not ready for commercialization, so are not eligible for registration at the relevant regulatory bodies. Second, the applied concept of "new sleep technologies" is broad, involving different types of technologies. Variations are expected on many aspects, including the intended public (clinical practitioners or general consumers), the format (hardware, software, etc.) and their purpose (diagnostic, therapeutical, follow-up, etc.), and regulation might vary for each of these conditions. Therefore, we prefer not to tie the definition of new sleep technologies to any regulatory aspect. In any case, regulation, approval or clearance by health regulation bodies are topics to be evaluated during the consensus voting. Thus, although not being part of the definition of new sleep technologies, recommendations related to legislation and approval by regulatory agencies are likely to be included in the resulting guidelines.

3 | FORMAT OF THE GUIDELINES

The presentation of items in this study will be organized on a three-level hierarchy, including categories (and subcategories when applicable), topics and practical items.

"Categories" refer to the categorization of sleep technologies that will be covered in this guideline. It aims to define the types of sleep technologies for which the guidelines will develop recommendations. The categories cannot be determined beforehand, as they depend on the nature of the responses during the core committee meeting and throughout the Delphi consensus process. In any case, it is likely that the categories will cover different domains related to new sleep technologies, including different sleep-related purposes (e.g. sleep monitoring, insomnia, obstructive sleep apnea, etc.), intended user (e.g. CST, medical-grade devices), presentation (e.g. wearable, nearable, bed sensor, app, etc.) and intended use (e.g. screening, diagnosis, follow-up/long-term monitoring and treatment). The level of specificity of these categories might vary, from broad and single-domain categories (e.g. recommendations applying to all sleep-related wearable devices, or to all technologies aiming at screening obstructive sleep apnea) to more specific subcategories combining two or more domains (e.g. wearable SpO_2-based sleep apnea devices).

"Topics" refer to each methodological topic that should be considered on the process of development, performance evaluation and validation of new sleep technologies. It is understood that the practices relating to each of these topics might vary, with some being considered more adequate to assure the functionality, reliability, safety and accuracy of sleep technology, while others are considered as substandard. Possible topics include but are not limited to sample size, research design, patient selection, diagnostic criteria, scoring strategy, control group characteristics, algorithm development, data availability, outcome detection tools, statistical analysis, hardware configuration, sampling rate, follow-up duration and registration on health regulatory agencies.

"Practical items" refer to different possible practices that might be performed within each topic. More than one practical item can be listed for each item. Ideally all possible variations on practice should be listed, ranging from the "worst" to the "best" possible practices, regardless of how often they are actually performed or implemented.

The practical items are the level for which recommendations will be made. In each practical item, a status of consensus and a recommendation level will be informed. Status of consensus discloses in which step during the Delphi consensus process the consensus regarding was reached. Possible responses are "Panelists—first assessment", "Panelists—second assessment", "Core committee—first assessment", "Core committee—second assessment" and "no consensus reached" (for more details, check the "Phase 2: Delphi surveys" subsection). There will be three possible recommendation levels: "unacceptable"; "acceptable"; and "ideal".

The checklist will list all the topics included in the guidelines, and should be employed to evaluate the group of articles, studies and experiments used to develop and validate new sleep technologies (rather than to evaluate the methodological quality of a single study).

The checklist will contain two fillable columns, one to indicate the highest recommendation level reached for a given topic and another to indicate in which article it has been described.

4 | DELPHI METHODOLOGY

A Delphi study is a systematized method used to reach consensus on a given topic based on the inputs and contributions from a group of specialists (Diamond et al., 2014; Jünger et al., 2017; McPherson et al., 2018). It has been extensively used for the development of standards or guidelines, and for orienting decision-making in healthcare, especially when empirical evidence is limited, controversial or not applicable (Barrett & Heale, 2020; Humphrey-Murto et al., 2020; Niederberger & Spranger, 2020). This method was originally developed by the RAND corporation in the context of reaching consensus and guiding decisions for strategic military questions (McPherson et al., 2018). In the following decades and especially since the 1970s, the Delphi method was adapted for use in biomedical and healthcare research (Barrett & Heale, 2020; McPherson et al., 2018). Since then, this method has been extensively utilized, including in sleep medicine (Berezin et al., 2021; Boerner et al., 2015; Drager et al., 2023; Murphie et al., 2018; Pearson et al., 2020; Studart-Pereira et al., 2023). However, there are no standard quality parameters or unified guidelines on how to perform it (Nasa et al., 2021), which leads to several modified versions and substantial variability on methodological aspects and reporting characteristics (Boulkedid et al., 2011; Diamond et al., 2014; Niederberger & Spranger, 2020; Spranger et al., 2022). In this study we will apply a modified Delphi methodology. The term "modified Delphi" has been used to describe studies that are based on the original Delphi method and that maintain its core characteristics, but that present some adaptations to the original protocol. The nature of these modifications varies considerably among protocols, in a way that there is not a standard form of a "modified Delphi" (some of them including the presence of face-to-face meetings, usage of different analytical methods, and the employment of multiple panels of specialists; Niederberger & Spranger, 2020). For that reason, it is important that any modified Delphi study provides a comprehensive description of the methods (Humphrey-Murto et al., 2017).

In general, Delphi studies are based on four core characteristics: anonymity; iteration; statistical response estimates; and controlled feedback (Humphrey-Murto et al., 2020; Nasa et al., 2021; Niederberger & Spranger, 2020).

- ***Anonymity***: Panellists participate anonymously in Delphi panels. Anonymity is essential, as it assures independent and unbiased participation.
- ***Iteration***: A Delphi panel is composed by iterative rounds, in which the propositions, answers and results of a previous round can be reassessed on next rounds.
- ***Statistical response estimates***: The responses are quantified, and agreement rates can be calculated to evaluate if consensus is achieved.
- ***Controlled feedback***: Participants receive their inputs back after each round, so it can be compared with the group decision and eventually be considered on further rounds.

The participants of this Delphi study will be distributed into three categories: steering committee; core committee; and panel of specialists. Specific attributions of each of these levels of participation are presented in Table 1. Further information regarding the composition of each participation level, the evaluation of conflicts of interests and the eligibility criteria to join the study is discussed in the supplementary material in Appendix S1. The practical steps on the performance of this study are detailed below.

4.1 | Phase 1: First core committee meeting

A first core committee meeting will be scheduled among all core committee members. This meeting will take place online and participation is mandatory for an author to be considered as a core committee member. This meeting is intended as an opportunity to discuss and clarify the aims and directions of this consensus study. It includes presenting and discussing the following topics: authorship criteria, schedule and deadlines. dissemination plans, and definition of categories of devices, technologies and topics to be included in the consensus.

The definition of categories and topics to be included in the consensus will be reached on an open propositive discussion among the core committee members, and each proposal will be voted on. Possible answers are "yes", "no" and "unsure". The consensus threshold will be set at 66% (unweighted), which means that for a certain category to be considered included, at least 2/3 of the valid answers should have been "yes".

A category should be understood as a specific context or group of technologies for which specific recommendations should be developed. They cannot be determined beforehand, as they depend on the nature of the proposals. Likewise, topics within each category cannot be defined beforehand. Possible topics include but are not limited to sample size, research design, patient selection, diagnostic criteria, scoring strategy, algorithm development, data availability, statistical analysis, hardware configuration, sampling rate, follow-up duration. The panel facilitator and the steering committee reserve the right to exclude suggestions that are considered out of the scope of this study.

4.2 | Phase 2: Delphi surveys

Once categories and topics were decided in the core committee group, a modified electronic Delphi consensus study will be performed. It will be composed of up to four rounds, alternating from rounds of proposal and rounds of consensus. The surveys will be prepared in Welphi, a platform specialized on online Delphi studies. Responses will be quasi-anonymous, meaning that the panellists will be anonymous among themselves, but that the facilitator will be able

TABLE 1 Levels of participations and responsibilities.

Levels of participation	Responsibilities	Conflict of interests	Number of participants
Panel facilitator	• Organize and manage meetings • Prepare the platform for proposals and consensus in each Delphi round • Collect, analyse and organize answers following each Delphi round	High-level conflict of interests not allowed	1
Steering committee	• Nominate and select core committee members • Shortlist and select researchers for the board of panellists • Oversee, analyse and approve the results of each Delphi round • Exclude items that are out of the scope of the guidelines • Organize the working groups for the development and writing of the final report	High-level conflict of interests not allowed	4
Core committee	• Propose categories of new sleep technologies and topics to be included on the consensus • Revise and aim to reach consensus for cases in which no consensus was reached by the board of panellists • Discuss and define authorship criteria, schedules and deadlines, and dissemination plans • Coordinate work groups for the development and writing the final report	Conflict of interests allowed	15
Panel of specialists	• Propose practical items and vote on the Delphi rounds	Conflict of interests allowed, but resulting in weighted responses during Delphi consensus	60–80

to identify the responses in order to manage information from one round to the other. Surveys in each round will be open for 14 days, and reminder e-mails will be sent at days 7, 10 and 14, to assure high responses rates. Finally, all the responses and results from all the iterations throughout the Delphi consensus process will be made available as supplementary materials in Appendix S1, in order to assure transparency on the process of achieving consensus.

On the proposal rounds (rounds 1 and 2), panel members will be able to propose items to be voted on following rounds. Proposing new items should follow the scope of this study. This is intentionally open, in order to not restrict panellists on their suggestions. It is natural and expected that some proposed items fall out of scope, but that is likely to be filtered throughout the process, first by the steering committee who holds the right to exclude items that are inappropriate in all rounds; second, by the process of achieving consensus, either by the panellists or by the core committee, which will naturally filter proposals that are evidently out of scope. The consensus rounds (rounds 2, 3 and 4) are intended to determine the levels of recommendation. These are the rounds in which panellists will have the possibility to evaluate the level of recommendation of each of the proposed items (within each of the topics and categories). For each item, the panellists will be able to determine if they consider it as "unacceptable", "acceptable" or "ideal". A detailed description of the procedures within each of the four Delphi rounds can be found in the supplementary material in Appendix S1.

The decision about each of these three answers is subjective, but should consider evidence level and clinical design, clinical reliability, confounding factors and risk of bias, patient/user safety and medical/research ethics. Factors related to commercial aspects, costs involved and time requirements should not be taken as factors for decision. Further instruction for the decision about each recommendation level is provided in the supplementary material in Appendix S1. Taking that into consideration, the use of recommendation levels can be understood as disclosed below (further information provided in the supplementary material in Appendix S1).

"Unacceptable" refers to practices that are considered not acceptable, unreliable or insufficient for the development, performance evaluation or validation of new sleep technologies. Considering a technique as unacceptable means that it should not be performed, or that if performed, contributes nothing or very little to the process of development, performance evaluation and validation of new sleep technologies.

"Acceptable" refers to techniques that are evidently not the best practices available, but that might be accepted on the process of development, performance evaluation and validation of a new sleep technology. Considering an item as acceptable means it is valid on this process, but also denotes that there are practices that might be more appropriate. It also means that there are better methods available for that given topic. As the process of development, performance evaluation or validation of new sleep technologies hardly involve a single study, "acceptable" items could be understood as intermediate steps on this process, until "ideal" practices are implemented.

"Ideal" refers to techniques that are the best possible practices for a given practical topic. Considering an item as "ideal" means that

the process of development, validation and performance evaluation of new sleep technologies should be using them whenever possible to truly assure its validity, accuracy and reliability.

The consensus threshold will be set at 66%, which means that a definitive recommendation level will be reached if at least 2/3 of weighted answers are given for the same answer. All the calculations will be weighted by the level of conflict of interests of each panellist. Voting of individuals reporting no risk of bias will be weighted as 1, while individuals reporting low risk of bias will be weighted as 0.5, and individuals reporting high risk of bias will be weighted as 0.25. A complete description of the calculation of consensus scores is available in the supplementary material in Appendix S1.

4.3 | Phase 3: Second core committee meeting

After all the Delphi rounds have been completed, the second core committee meeting will happen. This meeting will take place online, and participation is mandatory for an author to be considered as a core committee member.

This group meeting has three main objectives: (1) to revise all the recommendations reached by the panel members; (2) to discuss and reach consensus for items for which no consensus was reached by the panel members; and (3) to organize working groups to prepare the final guidelines.

All the recommendations will be revised at the core committee meeting. It is expected that all items for which consensus was reached by the panel members are endorsed by the core committee. However, the core committee holds the right to change the level of recommendation of any item, provided that there is a robust reason for that, and that there is an absolute consensus among all the core committee members participating in the meeting. Possible reasons to change a recommendation level include correcting logical misconceptions on the level of recommendations (such as less robust technique receiving a higher level of recommendation than a more robust one) and the need to include new technologies developed.

The core committee will also discuss and revise all the items for which no consensus was reached by the panel members (either by a draw or by inconsistent responses) on a three-step approach. First, the core committee members will vote if it is valid and important to reach consensus for each of these items. A simple majority of votes (>50%) is sufficient to conclude that the core committee member should try reaching consensus, then following the same two steps used to reach consensus by the panel members. As an important difference, all the consensus processes will take place synchronously and openly during the core committee meeting.

Finally, work groups will be determined for each category of sleep technologies for which recommendations were drawn. The work groups will be responsible for writing the recommendations based on the consensus reached by both the panel members and the core committee. Each work group will have a leader and might encompass two more individuals (either core members or panellists). Cases for which the core committee considered reaching a consensus was not relevant, or when no consensus was reached by the core committee will be labelled as "no consensus reached".

4.4 | Phase 4: Preparation of final recommendation and resulting manuscripts

Each work group will prepare a document containing the recommendations for each item, organized by categories and topics. The recommendations for each category will be presented in tables, in which the topics and their items will be listed. Within each topic, the items will be ranked by recommendation level, in the following order: ideal; acceptable; unacceptable; and no consensus reached. Also, the way the consensus was reached should be disclosed, as: "panel members—first round"; "panel members—second round"; "core committee—first round"; "core committee—second round"; or "no consensus reached". Finally, the agreement rate will be disclosed. For each item, the work groups shall prepare a short explanatory text, discussing its relevance, implementation and providing evidence when available.

All documents prepared by each work group will be sent to the steering committee, who will be responsible for merging the contributions and assure consistency on writing and language style. Based on that, the steering committee shall deliver two files: a final recommendations file and a checklist including all items. Once the final files are prepared, they will be submitted to all panellists for edits and suggestions. Up to two review rounds are allowed, and only grammar correction or alterations on text presentations are expected. No major changes are allowed, especially regarding including new items or changing recommendation levels.

4.5 | Phase 5: Final meeting

The final meeting intends to present the final files to all core committee members and panellists. Plans might change in case of delays in the intended schedule. This meeting has only an informative nature, and no voting or any other collaborative work is expected. Attending to this meeting is not an authorship criterion. Items to be discussed include dissemination, implementations, monitoring and updating plans.

5 | EXPECTED RESULTS AND USAGE OF THESE RECOMMENDATION GUIDELINES

The resulting recommendations, especially the checklist, are not intended to be used as a quality assessment tool to evaluate individual articles. The process of development and validation of a new sleep technology is complex, and it hardly involves a single study. More likely, multiple studies and experiments are needed in order for a given technology to be considered fully validated on all its aspects. Rather, it should be used to evaluate the overall procedures, studies and experiments performed to develop, evaluate performance and validate new technologies, either by a single company or by a group

of companies co-developing a technology. It can also be used by health regulatory authorities and agencies to evaluate devices and technologies that might apply for registration, approval, clearance or inclusion on health systems.

6 | CONCLUSION

As new sleep technologies are being developed, refined and made available, concerns regarding their accuracy and validation have been raised. The DevSleepTech guidelines is an effort to standardize these processes, covering new sleep technologies on the broadest way possible. We believe that standardizing the practices by which new technologies are developed will be beneficial for all stakeholders in the field, including: (1) to developers (technology companies and startups), who will have parameters to assure their technologies can be considered as validated, accurate and reliable; (2) to sleep researchers, who will have a clear guide on how to perform development and validation studies for new sleep technologies; (3) to medical professionals, who will be able to differentiate which sleep technologies are reliable and validated technologies based on objective criteria; (4) to patients and users, for whom the evidence and accuracy level of technologies that have been presented should increase.

AUTHOR CONTRIBUTIONS

Gabriel Natan Pires: Conceptualization; methodology; project administration; writing – original draft; writing – review and editing. **Erna S. Arnardóttir:** Conceptualization; funding acquisition; validation; writing – review and editing. **Sébastien Bailly:** Methodology; validation; writing – review and editing. **Walter T. McNicholas:** Conceptualization; validation; funding acquisition; writing – review and editing; supervision.

ACKNOWLEDGEMENTS

This work has received research funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 965417. GNP receives funding from the *Associação Fundo de Incentivo à Pesquisa* (AFIP). ESA reports additional funding from NordForsk (NordSleep project 90458) via the Icelandic Research Fund. SB is supported by the "e-health and integrated care and trajectories medicine" Chairs of excellence from the Grenoble Alpes University Foundation and MIAI artificial intelligence (ANR-19-P3IA-0003).

CONFLICT OF INTEREST STATEMENT

GNP is a shareholder at SleepUp© and founder of P&P Metanálises. ESA discloses lecture fees from Nox Medical, Jazz Pharmaceuticals, Linde Healthcare, Alcoa—Fjardaral, Controlant, Wink Sleep, Apnimed and Novo Nordisk (via Vistor). ESA is also a member of the Philips Sleep Medicine & Innovation Medical Advisory Board. The other authors have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Gabriel Natan Pires https://orcid.org/0000-0003-0411-0111
Erna S. Arnardóttir https://orcid.org/0000-0003-0877-3529
Sébastien Bailly https://orcid.org/0000-0002-2179-4650
Walter T. McNicholas https://orcid.org/0000-0001-5927-2738

REFERENCES

Baptista, P. M., Martin, F., Ross, H., O'Connor Reina, C., Plaza, G., & Casale, M. (2022). A systematic review of smartphone applications and devices for obstructive sleep apnea. *Brazilian Journal of Otorhinolaryngology*, *88*, S188–S197. https://doi.org/10.1016/j.bjorl.2022.01.004

Barrett, D., & Heale, R. (2020). What are Delphi studies? *Evidence-Based Nursing*, *23*(3), 68–69. https://doi.org/10.1136/ebnurs-2020-103303

Behar, J., Roebuck, A., Domingos, J. S., Gederi, E., & Clifford, G. D. (2013). A review of current sleep screening applications for smartphones. *Physiological Measurement*, *34*(7), R29–R46. https://doi.org/10.1088/0967-3334/34/7/R29

Benjafield, A. V., Ayas, N. T., Eastwood, P. R., Heinzer, R., Ip, M. S. M., Morrell, M. J., Nunez, C. M., Patel, S. R., Penzel, T., Pépin, J. L., Peppard, P. E., Sinha, S., Tufik, S., Valentine, K., & Malhotra, A. (2019). Estimation of the global prevalence and burden of obstructive sleep apnoea: A literature-based analysis. *The Lancet Respiratory Medicine*, *7*(8), 687–698. https://doi.org/10.1016/S2213-2600(19)30198-5

Berezin, L., Nagappa, M., Wong, J., Clivatti, J., Singh, M., Auckley, D., Charchaflieh, J. G., Jonsson Fagerlund, M., Gali, B., Joshi, G. P., Overdyk, F. J., Margarson, M., Mokhlesi, B., Moon, T., Ramachandran, S. K., Ryan, C. M., Schumann, R., Weingarten, T. N., Won, C. H. J., & Chung, F. (2021). Identification of sleep medicine and anesthesia Core topics for anesthesia residency: A modified Delphi technique survey. *Anesthesia and Analgesia*, *132*(5), 1223–1230. https://doi.org/10.1213/ANE.0000000000005446

Boerner, K. E., Coulombe, J. A., & Corkum, P. (2015). Core competencies for health professionals' training in pediatric behavioral sleep care: A Delphi study. *Behavioral Sleep Medicine*, *13*(4), 265–284. https://doi.org/10.1080/15402002.2013.874348

Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., & Alberti, C. (2011). Using and reporting the Delphi method for selecting healthcare quality indicators: A systematic review. *PLoS One*, *6*(6), e20476. https://doi.org/10.1371/journal.pone.0020476

de Zambotti, M., Cellini, N., Menghini, L., Sarlo, M., & Baker, F. C. (2020). Sensors capabilities, performance, and use of consumer sleep technology. *Sleep Medicine Clinics*, *15*(1), 1–30. https://doi.org/10.1016/j.jsmc.2019.11.003

de Zambotti, M., Menghini, L., Grandner, M. A., Redline, S., Zhang, Y., Wallace, M. L., & Buxton, O. M. (2022). Rigorous performance evaluation (previously, "validation") for informed use of new technologies for sleep health measurement. *Sleep Health*, *8*(3), 263–269. https://doi.org/10.1016/j.sleh.2022.02.006

Depner, C. M., Cheng, P. C., Devine, J. K., Khosla, S., de Zambotti, M., Robillard, R., Vakulin, A., & Drummond, S. P. A. (2020). Wearable technologies for developing sleep and circadian biomarkers: A summary of workshop discussions. *Sleep*, *43*(2), zsz254. https://doi.org/10.1093/sleep/zsz254

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*(4), 401–409. https://doi.org/10.1016/j.jclinepi.2013.12.002

Drager, L. F., Assis, M., Bacelar, A. F. R., Poyares, D. L. R., Conway, S. G., Pires, G. N., de Azevedo, A. P., Carissimi, A., Eckeli, A. L., Pentagna, Á., Almeida, C. M., & Martinez, S. C. G. (2023). 2023 guidelines on the diagnosis and treatment of insomnia in adults – Brazilian sleep association. *Sleep Science*, *16*(S02), 1–43.

Ferrie, J. E., Kumari, M., Salo, P., Singh-Manoux, A., & Kivimäki, M. (2011). Sleep epidemiology – a rapidly growing field. *International Journal of Epidemiology*, *40*(6), 1431–1437. https://doi.org/10.1093/ije/dyr203

Fino, E., & Mazzetti, M. (2019). Monitoring healthy and disturbed sleep through smartphone applications: A review of experimental evidence. *Sleep & Breathing*, *23*(1), 13–24. https://doi.org/10.1007/s11325-018-1661-3

Goldstein, C. A., & Depner, C. (2021). Miles to go before we sleep…a step toward transparent evaluation of consumer sleep tracking devices. *Sleep*, *44*(2), zsab020. https://doi.org/10.1093/sleep/zsab020

Humphrey-Murto, S., Varpio, L., Wood, T. J., Gonsalves, C., Ufholz, L. A., Mascioli, K., Wang, C., & Foth, T. (2017). The use of the Delphi and other consensus group methods in medical education research: A review. *Academic Medicine*, *92*(10), 1491–1498. https://doi.org/10.1097/ACM.0000000000001812

Humphrey-Murto, S., Wood, T. J., Gonsalves, C., Mascioli, K., & Varpio, L. (2020). The Delphi method. *Academic Medicine*, *95*(1), 168. https://doi.org/10.1097/ACM.0000000000002887

Jünger, S., Payne, S. A., Brine, J., Radbruch, L., & Brearley, S. G. (2017). Guidance on conducting and REporting DElphi studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliative Medicine*, *31*(8), 684–706. https://doi.org/10.1177/0269216317690685

Khosla, S., Deak, M. C., Gault, D., Goldstein, C. A., Hwang, D., Kwon, Y., O'Hearn, D., Schutte-Rodin, S., Yurcheshen, M., & Kirsch, D. B. (2019). Consumer sleep technologies: How to balance the promises of new technology with evidence-based medicine and clinical guidelines. *Journal of Clinical Sleep Medicine*, *15*(1), 163–165. https://doi.org/10.5664/jcsm.7598

Khosla, S., Deak, M. C., Gault, D., Goldstein, C. A., Hwang, D., Kwon, Y., O'Hearn, D., Schutte-Rodin, S., Yurcheshen, M., Rosen, I. M., & Kirsch, D. B. (2018). Consumer sleep technology: An American Academy of sleep medicine position statement. *Journal of Clinical Sleep Medicine*, *14*(5), 877–880. https://doi.org/10.5664/jcsm.7128

McPherson, S., Reese, C., & Wendler, M. C. (2018). Methodology update: Delphi studies. *Nursing Research*, *67*(5), 404–410. https://doi.org/10.1097/NNR.0000000000000297

Menghini, L., Cellini, N., Goldstone, A., Baker, F. C., & de Zambotti, M. (2021). A standardized framework for testing the performance of sleep-tracking technology: Step-by-step guidelines and open-source code. *Sleep*, *44*(2), zsaa170. https://doi.org/10.1093/sleep/zsaa170

Moher, D., Schulz, K. F., Simera, I., & Altman, D. G. (2010). Guidance for developers of health research reporting guidelines. *PLoS Medicine*, *7*(2), e1000217. https://doi.org/10.1371/journal.pmed.1000217

Murphie, P., Little, S., Paton, R., McKinstry, B., & Pinnock, H. (2018). Defining the Core components of a clinical review of people using continuous positive airway pressure therapy to treat obstructive sleep apnea: An international e-Delphi study. *Journal of Clinical Sleep Medicine*, *14*(10), 1679–1687. https://doi.org/10.5664/jcsm.7372

Nasa, P., Jain, R., & Juneja, D. (2021). Delphi methodology in healthcare research: How to decide its appropriateness. *World Journal of Methodology*, *11*(4), 116–129. https://doi.org/10.5662/wjm.v11.i4.116

Nester, R. (2019). Smart Sleep Tracking Device Market: Global Demand Analysis & Opportunity Outlook 2024.

Niederberger, M., & Spranger, J. (2020). Delphi technique in health sciences: A map. *Frontiers in Public Health*, *8*, 457. https://doi.org/10.3389/fpubh.2020.00457

Pearson, L., Parker, S., & Mansell, W. (2020). The development of a theoretically derived measure exploring extreme appraisals of sleep in bipolar disorder: A Delphi study with professionals. *Behavioural and Cognitive Psychotherapy*, *48*(4), 395–407. https://doi.org/10.1017/S1352465820000053

Penzel, T., Glos, M., & Fietze, I. (2021). New trends and new Technologies in Sleep Medicine: Expanding accessibility. *Sleep Medicine Clinics*, *16*(3), 475–483. https://doi.org/10.1016/j.jsmc.2021.05.010

Perez-Pozuelo, I., Zhai, B., Palotti, J., Mall, R., Aupetit, M., Garcia-Gomez, J. M., Taheri, S., Guan, Y., & Fernandez-Luque, L. (2020). The future of sleep health: A data-driven revolution in sleep science and medicine. *NPJ Digital Medicine*, *3*, 42. https://doi.org/10.1038/s41746-020-0244-4

Pires, G. N., Arnardóttir, E. S., Islind, A. S., Leppänen, T., & McNicholas, W. T. (2023). Consumer sleep technology for the screening of obstructive sleep apnea and snoring: Current status and a protocol for a systematic review and meta-analysis of diagnostic test accuracy. *Journal of Sleep Research*, *32*(4), e13819. https://doi.org/10.1111/jsr.13819

Santos-Silva, R., Bittencourt, L. R., Pires, M. L., de Mello, M. T., Taddei, J. A., Benedito-Silva, A. A., Benedito-Silva, A. A., Pompeia, C., & Tufik, S. (2010). Increasing trends of sleep complaints in the city of Sao Paulo, Brazil. *Sleep Medicine*, *11*(6), 520–524.S1389-9457(10)00156-5 [pii]. https://doi.org/10.1016/j.sleep.2009.12.011

Schutte-Rodin, S., Deak, M. C., Khosla, S., Goldstein, C. A., Yurcheshen, M., Chiang, A., Gault, D., Kern, J., O'Hearn, D., Ryals, S., Verma, N., Kirsch, D. B., Baron, K., Holfinger, S., Miller, J., Patel, R., Bhargava, S., & Ramar, K. (2021). Evaluating consumer and clinical sleep technologies: An American Academy of sleep medicine update. *Journal of Clinical Sleep Medicine*, *17*(11), 2275–2282. https://doi.org/10.5664/jcsm.9580

Spranger, J., Homberg, A., Sonnberger, M., & Niederberger, M. (2022). Reporting guidelines for Delphi techniques in health sciences: A methodological review. *Zeitschrift für Evidenz, Fortbildung Und Qualität Im Gesundheitswesen*, *172*, 1–11. https://doi.org/10.1016/j.zefq.2022.04.025

Studart-Pereira, L. M., Bianchini, E. M. G., Assis, M., Bussi, M. T., Corrêa, C. D. C., Cunha, T. C. A., Drager, L. F., Ieto, V., Lorenzi-Filho, G., de Luccas, G. R., Brasil, E. L., & Consensus Group on Sleep-focused Speech-Language-Hearing Sciences. (2023). Brazilian consensus on sleep-focused speech-language-hearing sciences – 2023 Brazilian sleep association. *Sleep Science*, *16*(S02), 1–18.

Tangudu, V., Afrin, K., Reddy, S., Deutz, N. E. P., Woltering, S., & Bukkapatnam, S. T. S. (2021). Toward standardizing the clinical testing protocols of point-of-care devices for obstructive sleep apnea diagnosis. *Sleep & Breathing*, *25*(2), 737–748. https://doi.org/10.1007/s11325-020-02171-5

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Pires, G. N., Arnardóttir, E. S., Bailly, S., & McNicholas, W. T. (2024). Guidelines for the development, performance evaluation and validation of new sleep technologies (DEVSleepTech guidelines) – a protocol for a Delphi consensus study. *Journal of Sleep Research*, e14163. https://doi.org/10.1111/jsr.14163