American Journal of Psychiatric Rehabilitation

Expert Opinion Article

What is a clinically relevant improvement on an 11-point single-item rating scale?

Emina Išerić[1], Pauline A. Hendriksen[1], Dana M. Dijkgraaf[1], Pantea Kiani[1], Joris C. Verster[1,2,3,*]

[1]Division of Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, 3584CG Utrecht, The Netherlands.
[2]Cognitive Neurophysiology, Department of Child and Adolescent Psychiatry, Faculty of Medicine, TU Dresden, D-01307 Dresden, Germany.
[3*]Centre for Mental Health and Brain Sciences, Swinburne University, Melbourne, VIC 3122, Australia.

*** Correspondence:** Joris C Verster, PhD
Division of Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, 3584CG Utrecht, The Netherlands. E-mail : j.c.verster@uu.nl

ORCID ID: Emina Išerić: 0009-0000-5263-2833
Pauline A. Hendriksen: 0000-0002-9030-4432
Dana Dijkgraaf: 0009-0002-2742-907X
Pantea Kiani 0000-0003-3073-9749
Joris C. Verster 0000-0002-6455-2096

Abstract
Single item rating scales, ranging from 0 to 10, are widely utilized in research and clinical practice to provide quick, reliable assessments of constructs such as mood, quality of life, and immune fitness. These scales offer several practical advantages, including ease of administration, reduced patient burden, and straightforward interpretability. This expert opinion explores the distinction between statistical significance and clinical relevance on 11-point single-item scales, with a focus on immune fitness, i.e. the body's capacity to mount an appropriate immune response to health challenges. Drawing on existing literature, including studies in pain and dietary interventions, it is proposed that a minimum difference of 1.5 points on the single-item immune fitness scale may represent a clinically relevant improvement. Nevertheless, the interpretation of changes on single-item scales depends on the concept being evaluated, its impact on daily life, and the absolute scores reported.

Keywords: Single-item assessment; clinical relevance; statistical significance; immune fitness; mood; quality of life

The importance of single-item rating scales

Single-item rating scales are gaining popularity in both research and clinical practice. In research settings – such as surveys and clinical trials – they offer a quick and global assessment of a given construct. In clinical practice, they can be used for screening, aiding diagnosis, monitoring disease progression or recovery over time, and assessing whether an intervention id effective.

Compared to longer, multiple-item scales, single-item ratings are quick to complete, easy to interpret, and require no calculations or recoding. Their brevity significantly reduces patient burden, while still providing meaningful insights. Over the past decade, several 11-point single-item rating scales have been developed and validated to assess mood [1,2] and quality of life [3]. Patients can rate the severity of different mood aspects such as stress, anxiety, depression, loneliness, and fatigue on scales ranging from 0 (absent) to 10 (extremely severe). Alternatively, quality of life or sleep quality can be rated on a scale ranging from 0 (very poor) to 10 (excellent). Beyond mood and well-being, single-item rating scales have also been developed to assess health-related constructs such as alcohol hangover severity [4], attaining a healthy diet [5], and immune fitness [6].

Immune fitness, defined as the capacity of the body to respond to health challenges such as infections, is crucial for maintaining health, resolving disease and improving overall quality of life [6,7]. The single-item immune fitness scale (see Figure 1) ranges from 0 (very poor) to 10 (excellent). Assessments can be conducted momentarily or retrospectively for a chosen time period (e.g., the past 4 weeks). A retrospective assessment may better capture transient health events (e.g., a 2-week period of illness within the past 4 weeks) that a single momentary rating might miss.

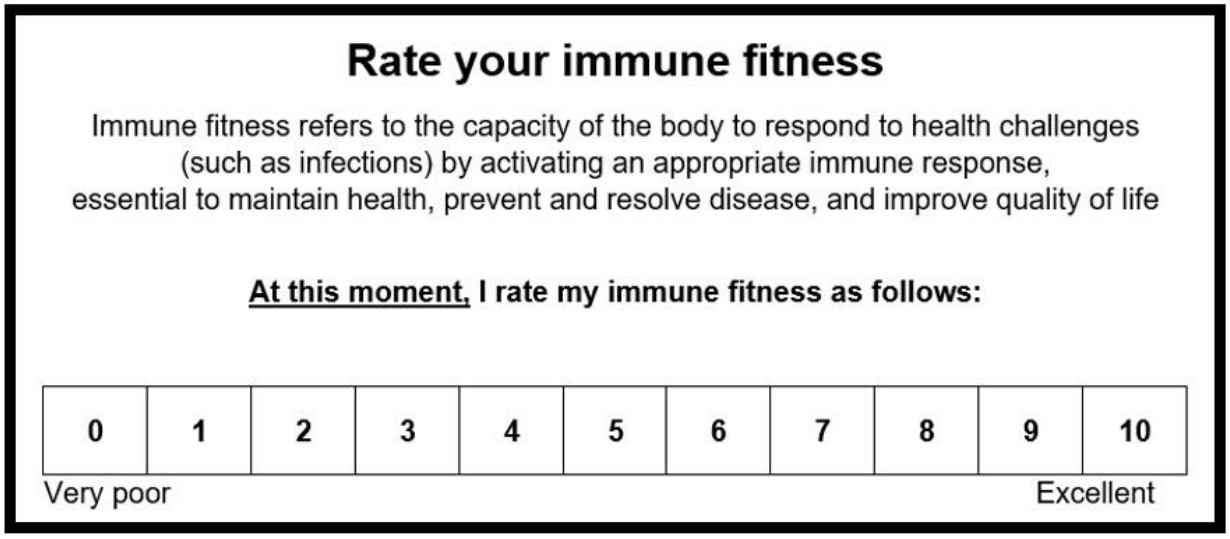

Figure 1. The single-item scale to assess immune fitness.

Beyond their practical advantages, single-item rating scales align with a growing emphasis on patient-centered care, where subjective experience is considered a key indicator of health. Verster et al. [1] demonstrated that single-item mood assessments significantly correlate with multi-item scales, while also capturing patients' holistic perception of their condition – something that more granular questionnaires may overlook. This holistic approach is particularly valuable in complex conditions such as immune-related disorders, where patients integrate multiple factors (e.g., fatigue, infection frequency, and recovery time) into a single rating. Immune fitness is hypothesized to be determined by various factors, including the type, number, frequency, duration, impact, and severity of immune-related complaints, as well as an individual's ability to cope with these complaints [7]. However, the relative contribution of each determinant to overall immune fitness remains unclear. Current multiple-item assessments of immune fitness primarily capture only a subset of these factors—typically type, number, and frequency. In contrast, a single-item assessment is presumed to integrate all relevant determinants, weighted appropriately, into one comprehensive evaluation. This makes single-item scales ecologically valid and highly applicable in real-world clinical settings. Moreover, their simplicity makes them ideally suited for integration into digital health platforms, such as mobile apps or wearable devices. This facilitates real-time health monitoring and enables frequent, unobtrusive assessments, which are crucial in the era of telemedicine and personalized healthcare.

Immune fitness, in particular, is of growing clinical interest. Many chronic illnesses (e.g., diabetes) and psychiatric and psychological disorders (e.g., depression) have been linked to immune dysfunction [8]. Additionally, many commonly reported symptoms in clinical practice – such as colds or flu-like complaints – are immune-related. [9]. Therefore, a quick, valid, and reliable tool for assessing immune fitness is essential. Previous research revealed that scores on the single-item immune fitness scale correlate significantly with various health outcomes (e.g., COVID-19 symptom severity), lifestyle factors (e.g., diet, sleep quality) mood (e.g., stress, anxiety), and quality of life [6]. Importantly, immune fitness assessments tend to remain stable over time in the absence of significant changes in health condition or lifestyle [10]. This stability is further supported by findings showing that seasonal differences in disease risk are not reflected in seasonal variability in immune fitness [10]. As such, the single-item immune fitness scale is ideal for evaluating the effects of interventions or treatments in clinical trials.

Advantages of single-item scales

Single-item rating scales provide a rapid, global assessment that incorporates the respondent's subjective evaluation of all facets relevant to the concept under evaluation. Rather than focusing on individual components, a single-item approach evaluates the overall constellation of the concept – encompassing the type and presence of symptoms, their severity, and impact on daily life. Accordingly, the US Food and Drug Administration (FDA) recognizes single-item assessments as valid patient-reported outcome measures (PROs) capable of measuring the effect of a treatment or other intervention [11].

One advantage of single-item scales is that they use a continuum for scoring, allowing respondents to provide nuanced feedback rather than forcing them

into limited, predefined categories. In this context, using clearly defined 'extreme' anchors such as 'very poor' or 'excellent' is helpful, as they identify a clear starting and ending point on the 0-10 scale [12]. In contrast, more ambiguous anchors such as 'normal' or 'satisfactory' should be avoided, as they may be interpreted anywhere along the continuum, reducing reliability [12].

Furthermore, while single-item scales provide a comprehensive, overarching assessment, multiple-item questionnaires typically isolate selected components (e.g., symptom frequency or severity) and often fail to capture the full construct, including its subjective implications (e.g., impact on daily functioning). Although many scale ranges are possible (e.g., 1-7, 0-100, 0-1000), comparative research across 10 countries revealed that a 0-10 scale is most valid and reliable. Larger scale ranges do not improve accuracy; instead, they tend to encourage rounding, as seen in common responses such as 25, 50, or 75 on a 0-100 scale [13].

An additional advantage of single-item scales lies in their adaptability across diverse populations and cultural contexts. Scherpenzeel and Saris [13] demonstrated that the simplicity and universal interpretability of the 0-10 scale reduce cultural biases that can comprise multi-item scales, which may rely on nuanced or culturally-specific language. This cross-cultural reliability is particularly relevant for assessing immune fitness, where global prevalence of immune-related challenges (e.g., infections) is accompanied by significant variability in perceived impact. For instance, Verster et al. [6] found consistent correlations between immune fitness scores and health outcomes across different cohorts, highlighting the ability of single-item scales to transcend demographic and geographic variability and enhance their utility in international clinical research.

Moreover, the brevity of single-item scales makes them especially useful as screening tools in resource-limited settings, where time and trained personnel may be scarce. Their simplicity allows healthcare providers to quickly identify individuals at risk of reduced immune fitness, prompting further investigation or immediate intervention. For example, a noticeable decline on a single-item immune fitness scale may prompt a more detailed follow-up in a primary care setting, leveraging the scale's sensitivity to meaningful change. This practical application underscores the potential of single-item scales to bridge the gap between clinical research and frontline clinical decision-making.

Taken together, single-item scales offer several important advantages over traditional multiple-item questionnaires. However, an key challenge remains: how to interpret the ratings provided, and how to meaningfully evaluate changes in scores on single-item scales.

Clinical relevance versus statical significance

Clinical relevance refers to whether a patient feels a clear improvement in health status, daily functioning, or quality of life following a treatment or intervention. While statistically significant differences are informative for evaluating the efficacy of an intervention, it is often more important is to determine whether an observed difference (e.g., between drug and placebo, or between pre- and post-intervention scores) holds clinical relevance. For example, a small difference in height between two groups (e.g., 1.73 versus 1.72 m) may reach statistical significance in a sufficiently large sample, yet such a 1 cm difference is clinically irrelevant for the patient's health. Similarly, changes on an 11-point scale may be statistically significant, but the key question is whether the observed change is also clinically meaningful. In other words, how much improvement on an 11-point scale is required for the change to be considered clinically relevant?

The minimum clinically significant difference on a single-item scale depends on the concept being assessed. Different concepts vary in their impact on daily life and health. For example, a 2 points increase on a headache scale could significantly impair work performance, whereas a similar increase on a thirst scale might be bothersome but unlikely to affect daily activities. However, research investigating minimum clinically significant differences on an 11-point scale remains limited.

In one study assessing pain intensity in emergency department patients, changes in pain scores on the 11-point scale were compared with responses on a 5-point categorical scale: "a lot more," "a little more," "about the same," "a little less," or "a lot less" pain [14]. The difference between "a little more" or "a little less" severe was used to define the minimum clinically significant difference, which corresponded to a score change of approximately 1.41 points on the 11-point scale [14]. Another study by Farrar et al. [15] combined data from 10 different placebo-controlled clinical trials involving 2724 chronic pain patients treated with pregabalin [15]. The researchers compared scores on the 11-point pain intensity scale with patient-reported outcomes on the Patient Global Impressions of Change (PGIC) scale. An example of such a scale, to be completed at the end of an intervention study, is given in Figure 2.

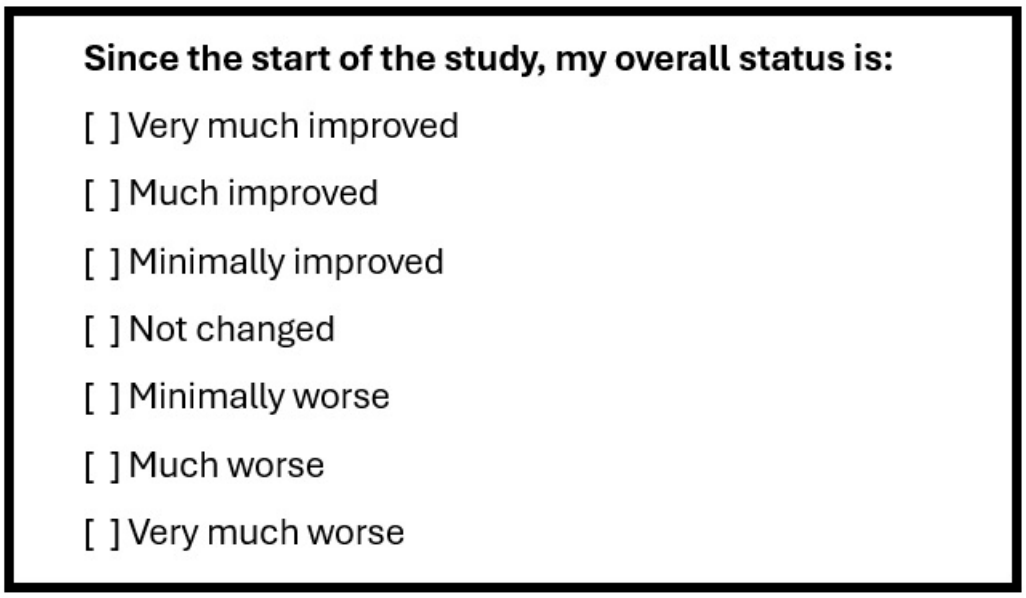

Figure 2. Example of a patient-reported global impressions of change (PGIC) scale.
Note: 'overall status' can be replaced by the concept under investigation (e.g., pain intensity or immune fitness).

Farrar et al. [15] found that a 2-point reduction on the 11-point scale corresponded to patients' ratings of "much improved" on the PGIC. Overall, anchor-based studies suggest that a clinically relevant improvement on an 11-point pain scale typically falls between 1 and 3 points, depending on factors such as baseline symptom severity, clinical context, and anchor stringency [16].

For single-item assessments of immune fitness, a minimum clinically significant difference has not been systematically established. Additionally, only a limited number of studies have applied the immune fitness scale in clinical trial settings, and most of these interventions did not yield statistically significant improvements [17]. Also, there is no alternative immune fitness questionnaire with clearly defined anchors (e.g., very poor, excellent) that could serve as a reference for determining clinical relevance. As such, further research is needed in this area.

One relevant study by Baars et al. [18] retrospectively examined the impact of a raw milk product diet among 390 Dutch adults (mean age of 54 years old), using the single-item immune fitness scale. Participants were classified into either a normal health group or a poor health group based on their baseline general health assessments. In the normal health group, immune fitness improved modestly but significantly (+0.5 for males +0.7 and females). These limited score changes may be attributed to a ceiling effect, as baseline immune fitness scores were already relatively high (8.0 for males; 7.6 and females), leaving little room for further improvement. In contrast, individuals in the poor health group reported greater score increases (+1.3 for males; +1.4 and females). However, Baars et al. did not explicitly evaluate the clinical relevance of these changes. Nonetheless, based on findings from both pain studies and this dietary intervention, a 1.5 point change on an 11-point single-item immune fitness scale is proposed as a clinically relevant difference.

In clinical practice, individual patients cannot report a 1.5 point change on the 11-point scale. Therefore, a 1- or 2-point change can be used as a practical guideline for identifying clinically relevant improvement at the individual level. In contrast, for clinical trials, the average change of 1.5 points can serve as a meaningful benchmark for group-level analyses. To support study design, a power analysis can estimate the minimum required sample size to detect a clinically relevant pre-post change of 1.5 points in a within-subject design [19].

Previous research involving n = 8064 participants who completed the single-item immune fitness scale reported a mean score of 7.4 with a standard deviation (SD) of 1.7 [20]. For conservative estimation in smaller trials, an SD of 2.0 is recommended. Based on this assumption, a sample size of n = 16 is needed to detect a statistically significant difference (α = 0.05) with 85% power. For 90% or 95% power, the required sample size increases to n = 19 and n = 23, respectively.

Are all difference scores the same?

As stated previously, the minimum clinically significant difference on a single-item scale depends on the concept being evaluated. Therefore, even if identical difference scores are observed across various constructs, their impact on daily life or health may differ. In addition, interpreting a

difference score should always be done in conjunction with the absolute score reported by the patient. The absolute value on a single-item rating scale can significantly influence how the difference is understood in clinical terms.

For instance, an observed change in immune fitness of 1.5 points may be critical for one patient's condition, yet relatively inconsequential for another. A 2-point reduction in immune fitness from 4 to 2, for example, may represent a transition from manageable illness to a level requiring hospitalization. In contrast, a reduction from 9 to 7 – although numerically identical – may hold minimal clinical relevance for one's health status, as both values fall within a range associated with adequate immune fitness. Similarly, a score increase from 4 to 6 suggests a clinically relevant improvement from reduced to adequate immune fitness. However, even this interpretation should be made with caution, as not all individuals consider a score below 6 to reflect compromised immune fitness [6].

To further illustrate this point, consider the hypothetical clinical scenario of Patient X, who has a baseline immune fitness score of 3, and reported having frequent infections. Following a nutritional intervention, Patient X's score improves to 5. Although this 2-point increase exceeds the proposed minimum clinically relevant difference of 1.5, Patient X may still experience minor infections or significant functional impairments. In other words, a clinically relevant improvement has occurred, but does not automatically equate to full recovery. This nuanced relationship between clinically relevant change scores and absolute scores highlights the need for tailored interpretive guidelines, ideally supported by cohort studies that correlate immune fitness ratings with objective health markers, such as infection rates or inflammatory biomarkers.

In conclusion, 11-point single-item scales are cost-effective, valid, and reliable PROs that can provide a global assessment of a target concept. While a 1.5-point change is suggested as a clinically relevant difference on the single-item immune fitness scale, interpretation should always account for the absolute scores and their clinical context.

Looking forward, broader adoption of single-item scales – such as the immune fitness measure – would benefit from the development of standardized protocols for their use in both clinical trials and practice. Such protocols could include guidance for integrating single-item scales into electronic health records, clinician training on their interpretation, and the establishment of normative data across diverse populations. These steps would help cement single-item scales as central tools for patient-reported outcome assessment, bridging the gap between research and real-world healthcare delivery.

Conflicts of Interest: Over the past 3 years, J.V. has received research support from Danone and Inbiose and acted as a consultant/advisor for Eisai, KNMP, Med Solutions, Mozand, Red Bull, Sen-Jam Pharmaceutical, and Toast!. J.V. owns stock from Sen-Jam Pharmaceutical. D.D., E.I. and J.V. received travel support from Sen-Jam Pharmaceutical. P.K. is CEO of PanGenix. P.H. has nothing to declare.

Funding: none.

References

1. Verster JC, Sandalova E, Garssen J, Bruce G. The use of single-item ratings versus traditional multiple-item questionnaires to assess mood and health. European Journal of Investigation in Health Psychology and Education 2021, 11, 183-198. https://doi.org/10.3390/ejihpe11010015
2. Verster JC, Mulder KEW, Hendriksen PA, Verheul MCE, van Oostrom EC, Scholey A, Garssen J. Test-retest reliability of single-item assessments of immune fitness, mood and quality of life. Heliyon 2023, 9, e15280. https://doi.org/10.1016/j.heliyon.2023.e15280
3. Verster JC, Išerić E, Ulijn GA, Oskam SMP, Garssen J. Single-item assessment of Quality of Life: scale validation and associations with mood and health outcomes. Journal of Clinical Medicine 2024, 13, 5217. https://doi.org/10.3390/jcm13175217
4. Verster JC, van de Loo AJAE, Benson S, Scholey A, Stock A-K. The assessment of overall hangover severity. Journal of Clinical Medicine 2020, 9, 786. https://doi.org/10.3390/jcm9030786
5. Van Oostrom EC, Hendriksen PA, van der Weij BRC, Boogaard AS, Mulder KEW, Verheul MCE, Tan S, Kraneveld AD, Vlieg-Boerstra B, Garssen J, Verster JC. The single-item Healthy Diet Scale: Validity and reliability. Food Research 2025, 9, 101-110. https://doi.org/10.26656/fr.2017.9(2).194
6. Verster JC, Kraneveld AD, Garssen J. The assessment of immune fitness. Journal of Clinical Medicine 2022, 12, 22. https://doi.org/10.3390/jcm12010022
7. Verster JC, Išerić E, Garssen J. The concept and assessment of immune fitness. Exploration of Immunology 2023, 3, 500-505. https://doi.org/10.37349/ei.2023.00116
8. GBD 2019 IMID Collaborators. Global, regional, and national incidence of six major immune-mediated inflammatory diseases: findings from the global burden of disease study 2019. EClinicalMedicine 2023, 64, 102193. https://doi.org/10.1016/j.eclinm.2023.102193
9. Finley CR, Chan DS, Garrison S, Korownyk C, Kolber MR, Campbell S, Eurich DT, Lindblad AJ, Vandermeer B, Allan GM. What are the most common conditions in primary care? Systematic

review. Canadian Family Physician 2018, 64, 832-840.

10. Išerić E, Dijkgraaf DM, Kiani P, Devenney LE, Benson S, Scholey A, Garssen J, Verster JC. Seasonal differences in disease risk are not reflected in seasonal variability in immune fitness. South Eastern European Journal of Public Health 2025, 26, 3760-3768. https://www.seejph.com/index.php/seejph/article/view/5831
11. Food and Drug Administration (FDA). Guidance for Industry. Patient-Reported Outcome Measures: Use in medical product development to support labeling claims. December 2009. https://www.fda.gov/media/77832/download
12. Saris WE, De Rooij K. What kind of terms should be used for reference points? In: Saris, W.E. (ed.). Variation in response functions: A source of measurement error in attitude research. Amsterdam, Sociometric Research Foundation, 1988.
13. Scherpenzeel AC, Saris WE. The validity and reliability of survey questions: A meta-analysis of MTMM Studies. Sociological Methods & Research 1997, 25, 341-383. https://doi.org/10.1177/0049124197025003
14. Kendrick DB, Strout TD. The minimum clinically significant difference in patient-assigned numeric scores for pain. American Journal of Emergency Medicine 2005, 23, 828-832. https://doi.org/10.1016/j.ajem.2005.07.009
15. Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001, 94, 149-158. https://doi.org/10.1016/S0304-3959(01)00349-9
16. Gupta A, Hochman ME, Gutman EN. Clinical importance of changes on a numerical pain rating scale'. In: Gupta A, Hochman ME, Gutman EN (eds). 50 Studies every anesthesiologist should know, fifty studies every doctor should know. New York, USA, 2018, Oxford Academic. https://doi.org/10.1093/med/9780190237691.003.0035
17. Nesari T, Kadam S, Vyas M, Huddar VG, Prajapati PK, Rajagopala M, More A, Rajagopala SK, Bhatted SK, Yadav RK, Ma-hanta V, Mandal SK, Mahto RR, Kajaria D, Sherkhane R, Bavalatti N, Kundal P, Dharmarajan P, Bhojani M, Bhide B, Harti SK, Mahapatra AK, Tagade U, Ruknuddin G, Venkatramana Sharma AP, Rai S, Ghildiyal S, Yadav PR, Sandrepogu J, Deogade M, Pathak P, Kapoor A, Kumar A, Saini H, Tripathi R. Ayuraksha, a prophylactic Ayurvedic immunity boosting kit reducing positivity percentage of IgG COVID-19 among frontline Indian Delhi police personnel: A non-randomized controlled intervention trial. Frontiers in Public Health 2022, 10, 920126. https://doi.org/10.3389/fpubh.2022.920126
18. Baars T, Berge C, Garssen J, Verster JC. The impact of raw fermented milk products on perceived health and mood among Dutch adults. Nutrition and Food Science 2019, 49, 1195-1206. https://doi.org/10.1108/NFS-12-2018-0347
19. ClinCalc LLC. Sample size calculator. Available at: https://clincalc.com/stats/samplesize.aspx (assessed 21 March 2025).
20. Kiani P, Mulder KEW, Balikji J, Kraneveld AD, Garssen J, Verster JC. Pandemic preparedness: maintaining adequate immune fitness by attaining a normal, healthy bodyweight. Journal of Clinical Medicine 2022, 11, 3933. https://doi.org/10.3390/jcm11143933