UCSF

UC San Francisco Previously Published Works

Title

A Delphi-method-based consensus guideline for definition of treatment-resistant depression for clinical trials

Permalink

https://escholarship.org/uc/item/6g73b5m8

Journal

Molecular Psychiatry, 27(3)

ISSN

1359-4184

Authors

Sforzini, Luca
Worrell, Courtney
Kose, Melisa
et al.

Publication Date

2022-03-01

DOI

10.1038/s41380-021-01381-x

Peer reviewed

REVIEW ARTICLE   OPEN

Check for updates

A Delphi-method-based consensus guideline for definition of treatment-resistant depression for clinical trials

Luca Sforzini [1✉], Courtney Worrell [1], Melisa Kose[1], Ian M. Anderson[2], Bruno Aouizerate[3], Volker Arolt[4,5], Michael Bauer [6], Bernhard T. Baune[4,7], Pierre Blier[8,9,10], Anthony J. Cleare [1,11], Philip J. Cowen [12], Timothy G. Dinan [13], Andrea Fagiolini[14], I. Nicol Ferrier[15], Ulrich Hegerl[16], Andrew D. Krystal [17,18], Marion Leboyer [19], R. Hamish McAllister-Williams[15,20], Roger S. McIntyre[21,22,23,24], Andreas Meyer-Lindenberg[25], Andrew H. Miller [26], Charles B. Nemeroff [27], Claus Normann [28], David Nutt [29], Stefano Pallanti[30], Luca Pani[31,32,33], Brenda W. J. H. Penninx[34], Alan F. Schatzberg[35], Richard C. Shelton[36], Lakshmi N. Yatham[37], Allan H. Young [1,11], Roland Zahn[1,11], Georgios Aislaitner[38], Florence Butlen-Ducuing[39], Christine Fletcher[40], Marion Haberkamp[38], Thomas Laughren[41], Fanni-Laura Mäntylä[42], Koen Schruers[43,44], Andrew Thomson[45], Gara Arteaga-Henríquez[46,47,48], Francesco Benedetti [49,50], Lucinda Cash-Gibson[51], Woo Ri Chae [52], Heidi De Smedt[53], Stefan M. Gold [52,54,55], Witte J. G. Hoogendijk[56], Valeria Jordán Mondragón[57], Eduard Maron[58,59,60], Jadwiga Martynowicz[61], Elisa Melloni[49,50], Christian Otte [52], Gabriela Perez-Fuentes[62,63,64], Sara Poletti [50], Mark E. Schmidt[65], Edwin van de Ketterij [66], Katherine Woo[61], Yanina Flossbach[67], J. Antoni Ramos-Quiroga [46,48,62,64], Adam J. Savitz[68] and Carmine M. Pariante [1,11]

Criteria for treatment-resistant depression (TRD) and partially responsive depression (PRD) as subtypes of major depressive disorder (MDD) are not unequivocally defined. In the present document we used a Delphi-method-based consensus approach to define TRD and PRD and to serve as operational criteria for future clinical studies, especially if conducted for regulatory purposes. We reviewed the literature and brought together a group of international experts (including clinicians, academics, researchers, employees of pharmaceutical companies, regulatory bodies representatives, and one person with lived experience) to evaluate the state-of-the-art and main controversies regarding the current classification. We then provided recommendations on how to design clinical trials, and on how to guide research in unmet needs and knowledge gaps. This report will feed into one of the main objectives of the EUropean Patient-cEntric clinicAl tRial pLatforms, Innovative Medicines Initiative (EU-PEARL, IMI) MDD project, to design a protocol for platform trials of new medications for TRD/PRD.

INTRODUCTION

Many medications have proven efficacy in major depressive disorder (MDD) [1, 2], but frequently and even with multiple medication exposures, they fail to improve MDD symptoms [3–7]; one third of individuals do not achieve full symptomatic remission [3], and even fewer meet criteria for both symptomatic and functional remission [8]. In individuals with ineffective initial treatments, even if subsequent treatments are effective, there is a very high relapse rate while continuing the treatment; for example, in the STAR*D trial, individuals who required more treatment steps had higher relapse rates (up to 71% after the fourth step) [3].

'Incomplete response' is not an all-or-nothing phenomenon, but rather a continuum that ranges from partially responsive depression (PRD), to treatment-resistant depression (TRD), to 'multi-therapy-resistant MDD (MTR-MDD)' [9], to 'refractory depression', which implies an absence of response to all currently available treatments. Unfortunately, there is a lack of consensus definitions around concepts such as PRD, TRD, and 'adequate' treatments [10–13]. Moreover, guidelines on how to treat TRD/PRD, such as pharmacological augmentation, are not consistent [14], and evidence on

effectiveness is sparse [15]. This complicates the generalizability of results from research settings to the real-world, and hinders progress in this field, as there is no uniform population for clinical and biological investigations in TRD/PRD, including clinical trials for new or repurposed medications. Importantly, regulators acknowledge that response, partial response, and non-response exist on a continuum without universally accepted definitions, but nevertheless distinguish between these conditions; indeed, treatments for TRD and PRD already have accepted regulatory paths for drug approval.

Interestingly, a recent consensus statement (including experts also participating in the present report) suggested that the terms PRD and TRD are semantically and operationally not ideal, and proposed the broader concepts of 'difficult-to-treat depression (DTD)' or 'suspected DTD' [7], described as "depression that continues to cause significant burden despite usual treatment efforts". This concept overlaps with PRD and TRD, but introduces a more flexible, multidimensional and longitudinal definition. The authors themselves acknowledged that "what constitutes significant burden" is "subjective and likely to vary between patients", thus implicitly involving the patient's point of view, and also likely to vary

among clinicians and raters. Indeed, the authors cautioned that this definition may not be specific and objective enough to define clinical populations for regulatory clinical trials. In a subsequent follow-up paper, they also emphasised this as a 'model of care' that might be useful for both individuals with MDD and clinicians, especially in averting the development of 'therapeutic nihilism' [16].

In this document, we use a Delphi-method-based consensus approach to define TRD and PRD and to deliver operational criteria for future clinical studies, including clinical trials conducted for regulatory purposes. We have reviewed the relevant literature and brought together a group of international experts (including clinicians, academics, researchers, employees of pharmaceutical companies, regulatory bodies representatives, and one person with lived experience (PWLE)) to discuss the current state-of-the-art and the main controversies regarding the current classification. Our specific aims are to: (1) deliver TRD/PRD definitions that could be used at a person-centred level using information that is currently routinely collected in clinical practice; (2) recommend measures and instruments that should be included in future investigations and clinical trials for TRD/PRD; and (3) indicate which are the important areas that require further research. We provide consensus recommendations and the level of agreement for each recommendation, and discuss their limitations. We balance the need of clinicians and scientists with those of regulatory authorities, and describe differences when indicated.

This initiative is part of the EU Patient-cEntric clinicAl tRial pLatforms (EU-PEARL) programme, a public-private strategic partnership funded by the Innovative Medicines Initiative (IMI) to conceptualize and lead the design of an integrated research platform (IRP), that is, an infrastructure which allows the planning and completion of platform trials (see Appendix for further information).

OUR DELPHI-METHOD-BASED APPROACH
The Delphi approach is a method of choice for developing guidelines in health research [17]. Indeed, a recent commentary to the aforementioned paper on DTD [7] specifically recommends using a Delphi approach for consensus statement in depression [18]. However, this method could be difficult to standardize, with poor reporting of findings [19, 20]. Probably the most important of these issues has been identified in the definition of consensus [21]. Therefore, we clearly defined consensus as a percentage of agreement (vs. disagreement) on a precise recommendation, and further defined strong consensus when this percentage was equal or above 95%, moderate consensus between 61 and 94%, and weak consensus between 51 and 60%. Details of our Delphi process are available in the Appendix and summarized in Fig. 1. Briefly, it consisted of: (1) identification of experts with a track record of publications in this area or stakeholders with clear expertise, from clinical practice, academia, industry, and regulatory agencies; (2) first draft report, with an up-to-date narrative review on TRD/PRD definitions (in February 2020) and a questionnaire to gather opinions on the debated issues; (3) an online consensus meeting with academics and clinicians, on the 22nd of May 2020; (4) second draft report, integrating comments to the first draft with those from the meeting, and again circulated to all contributors; (5) third draft report (with an updated narrative review, in September 2020), circulated to a different group of stakeholders, including representatives of regulatory authorities, industry, and one PWLE; (6) an online consensus meeting with stakeholders, on the 9th of October 2020, with additional stakeholders providing written feedback; (7) fourth draft report (with a systematic review of the literature published from March 2020 to January 2021), integrating stakeholders comments, which was then circulated and approved by all authors, and submitted in its entirety (approximately 28,000 words) internally to IMI EU-PEARL (available from the corresponding author on request); (8) editing of the report into this shorter version for publication, again circulated and approved by all authors. Comments to the different

versions of the report were always anonymized. Methodological details of the narrative reviews and of the subsequent systematic review are presented in the Appendix.

We provide a number of recommendations, each based on the view supported by the largest number of the experts, while highlighting areas of uncertainty and defining the 'level of consensus' on each recommendation: 'strong', indicating the top one-third of recommendations that had unanimous or almost unanimous (≥95%) agreement; 'moderate', indicating almost all the remaining two-thirds of recommendations, which had a substantial majority (61–94%); and 'weak', indicating the one recommendation that only reached the minimum consensus of 51%. We have summarised all the recommendations in Table 1 and Fig. 2. We suggest that: (1) recommendations with strong consensus should be immediately adopted; (2) recommendations with a moderate consensus could provide a working model for immediate implementation, but should be further confirmed by future research/discussion; and (3) recommendations with a weak consensus represent an area of uncertainty and thus require further research/discussion.

TRD AND PRD DEFINITIONS
What is lack of response?
A treatment response in MDD is defined by a reduction of at least 50% in MDD severity [22] on a standardised rating scale (see below); 'lack of response', encompassing both TRD and PRD, is the mirror image of this definition, that is, less than 50% reduction in depression severity. TRD is further associated, in most definitions, with the 'classic criteria' of lack of response to at least two medications at an adequate dose and duration (see below). There is, however, much variability around the definition of TRD. In a recent systematic review, Brown et al. [12] reported that, out of 155 TRD definitions identified in the published literature, 48.4% specified at least two sequential treatment failures as a requirement. In an even more recent systematic review, Gaynes and colleagues [5] found that only 37% of intervention studies in TRD had enrolled individuals with MDD meeting the criteria of at least two failed antidepressants, and only 19% had also described failure to adequate doses and durations of treatments. In fact, the most common definition for TRD in intervention trials involved a minimum of only one previous failed treatment (48%). Thus, the majority of studies on TRD do not seem to use the 'classic criteria', making it difficult to pool or compare data across these TRD studies. Not surprisingly, there was a strong consensus (98%) for our first recommendation that a definition of TRD for clinical trials conducted for regulatory purposes is necessary.

Operational criteria for TRD and PRD
Based on a number of guidelines and other expert documents [22–25], in this report we adopt the definition of TRD as indicating individuals who show a reduction of less than 25% in MDD severity to at least two antidepressants, and of PRD as indicating individuals who show a reduction of between 25% and <50% in MDD severity to at least one antidepressant. This should ideally be established using prospective psychometric assessments, or at least using clinical interviews and health records to measure retrospectively the improvement in depression severity (see below).

The majority of experts, even though with moderate consensus (85%), recommended the importance of this distinction between TRD and PRD for randomized clinical trials for new treatments, especially because of the potential advantage of separating these individuals for different types of randomized controlled trials (for example, switching for TRD vs. augmentation for PRD).

PREVIOUS ANTIDEPRESSANT TREATMENTS
Here we discuss: (1) how many previous treatments should be considered, (2) in which episode (current and/or past) and
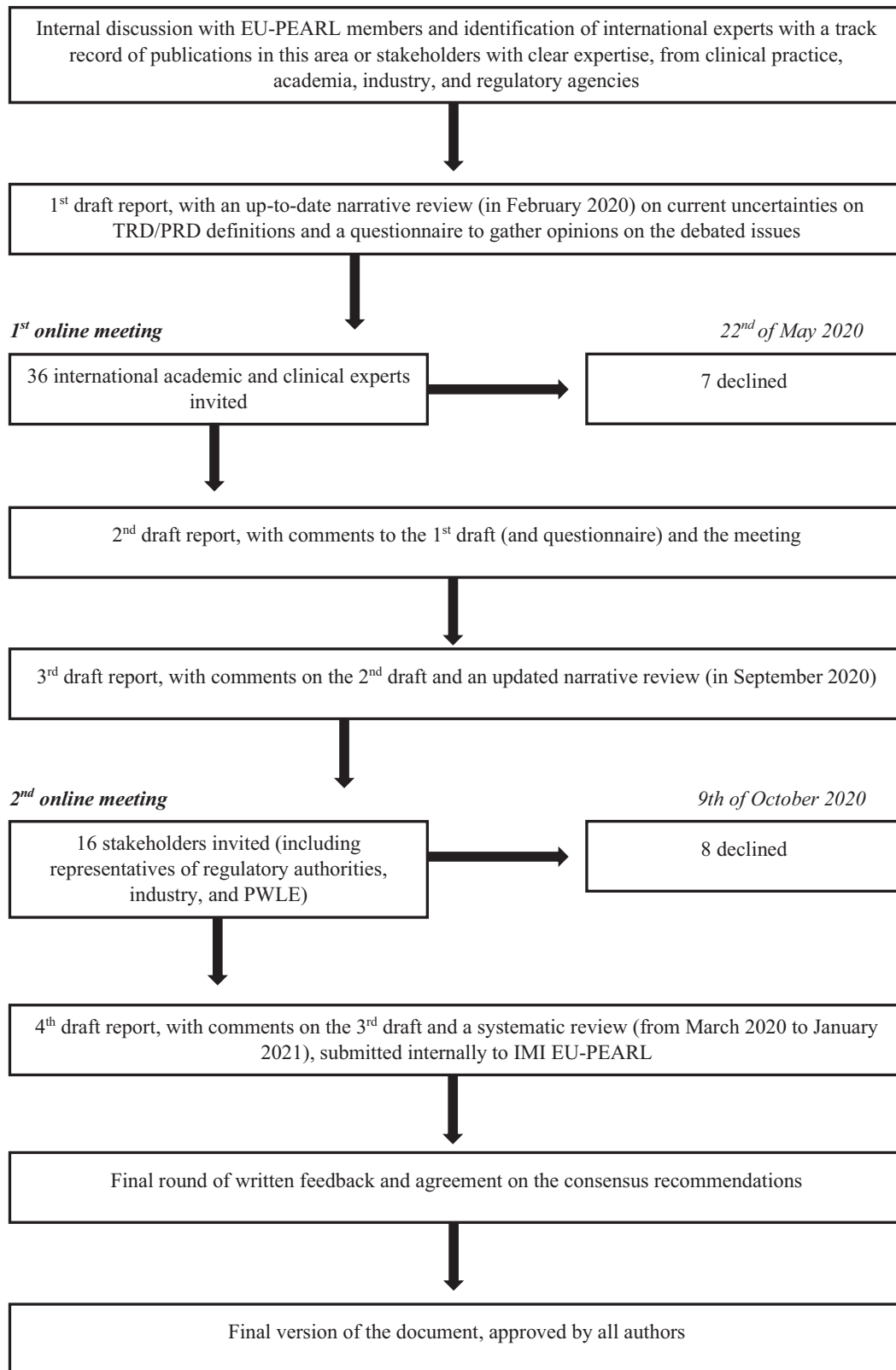
Fig. 1 **Delphi process flow diagram.** Schematic representation of the different phases of the Delphi-method-based process.

(3) the way the treatment failure or partial response should be assessed for inclusion in clinical trials (prospectively and/or retrospectively).

Number of previous treatments
The FDA guidance [26] acknowledges that no universally accepted definitions exist for TRD or PRD, and proposes that TRD studies

Table 1. Main consensus recommendations on TRD/PRD regulatory clinical trials.

Level of consensus– Strong		% of agreement
Recommendations which can be implemented within current practice		
TRD and PRD definitions		
1	A definition of TRD for clinical trials conducted for regulatory purposes is necessary.	98%
Previous antidepressant treatments		
3	TRD should be defined after a minimum of two failed treatments with <25% of improvement with adequate dosing and duration.	96%
Type of medications		
10	Discontinuation of treatment before the completion of the fourth week, without clear evidence of lack of response, should not be considered as a treatment failure for the purpose of establishing TRD/PRD.	96%
Exclusion from TRD/PRD studies		
12	A previous structured psychotherapy failing to improve MDD symptoms is not an exclusion criterion.	100%
Clinical presentation		
15	All specifiers of depression (melancholic, atypical, anxious, psychotic, mixed) should be considered within the TRD/PRD definition, except for bipolar depression.	95%
Recommendations which can be implemented in future research		
21	Future research should recognize and target different clinical phenotypes of TRD (and PRD) underpinned by a specific biological mechanism.	100%
22	For future research, diagnostic and history-taking instruments should be implemented in clinical cohorts and electronic health records, to allow a reliable, comprehensive, and multidimensional evaluation of people with lived experience (PWLE).	100%
24	Preferences, perspectives, and reported outcomes of PWLE should be included in future TRD (and PRD) diagnostic tools and outcome measures.	100%
Level of consensus – Moderate		
Recommendations which can be implemented within current practice		
TRD and PRD definitions		
2	It is important to distinguish between TRD and PRD for randomized clinical trials for new treatments.	85%
Previous antidepressant treatments		
4	PRD can be defined even after a single treatment (improvement 25 to <50%) with adequate dosing and duration.	76%
6	It is possible to assess ineffective past/current antidepressant treatment attempts, but only if properly documented, that is, based not only on subjective recollection or standardised instruments to assess psychiatric history and previous treatments (see below), but also on clinical documentation.	75%
Type of medications		
7	To define TRD, the two antidepressant treatment failures should consist of two established (licensed) medications for MDD of different mechanisms of action.	75%
8	A failed course of psychotherapy should not be included as one of the previous treatments required for the definition of TRD/PRD.	78%
9	The criteria of 'adequate dose and duration' is the minimal effective dosage, that is, the minimal approved dosage, administered for at least four weeks.	74%
Exclusion from TRD/PRD studies		
11	Multiple-drug resistant individuals, and individuals in whom augmentation strategies failed to improve/eliminate MDD symptoms should not be excluded from TRD/PRD studies.	93%
13	Individuals with MDD in whom deep brain stimulation (DBS) and vagus nerve stimulation (VNS) failed to improve/eliminate MDD symptoms should be excluded from TRD/PRD clinical studies.	62%
14	Individuals with MDD in whom other non-continuous/non-invasive brain stimulation interventions, such as electroconvulsive therapy (ECT) or transcranial magnetic stimulation (TMS), failed to improve/eliminate MDD symptoms should not be excluded.	88%
Clinical presentation		
16	Comorbid personality disorders or other mental disorders should be excluded from TRD/PRD studies only when their onset is properly documented as independent and antecedent to the MDD diagnosis.	79%
17	Individuals with a severe substance use disorder not currently in remission should be excluded from TRD/PRD studies, independently from the onset; in contrast, individuals with comorbid substance use disorder that is active and mild/moderate should be excluded from TRD/PRD studies only when the onset is properly documented as independent and antecedent to the MDD diagnosis.	83%
Diagnostic tools and measures of outcome		
18	Maudsley Staging Model is the preferred instrument to assess TRD/PRD status.	69%
19	Clinician administered MADRS10 is the preferred outcome instrument to assess treatment response (and remission), together with patient-reported QIDS-SR.	MADRS = 85% QIDS-SR = 81%
20	Criteria for remission, response, and partial response should not be relaxed in regulatory clinical trials for TRD/PRD, and shorter versions of the traditional scales, such as the HAM-D6 and the MADRS6, should not be preferred to full scales.	92%

Table 1 continued

	Recommendations which can be implemented in future research	
23	Currently, no biomarker has been validated in clinical practice or in clinical trials to identify people with TRD (and PRD), or to further stratify them; however, collection of biological samples for subsequent subgroup or stratified analyses is recommended.	91%
25	The usefulness of adherence assessment using blood levels or other methods (also in a run-in period) should be assessed through research, before deciding whether it should be implemented in future clinical trials.	92%
Level of consensus – Weak		
Recommendations which can be implemented within current practice		
Previous antidepressant treatments		
5	The definition of TRD should include two treatment failures both within the current episode, and the definition of PRD should include partial response to at least one treatment within the current episode; moreover, for long current episodes, only treatment failures within the last two years should be considered.	51%

should include people with MDD who have not responded to more than one (so, at least two) prior medications administered at an adequate dose and duration. Similarly, the EMA opts for the same definition of at least two failed treatments when considering the matter "in a clinical pragmatic view". This same definition is confirmed as the most common in the review by Gaynes and colleagues [5], and endorsed by the DTD consensus statement by McAllister–Williams and colleagues [7], although the last document states that some individuals might be considered to have DTD even with a single treatment failure, for example when standard treatments are contraindicated. There is also an uncertainty around PRD. EMA documents do not specify a precise number of previous treatments to diagnose PRD [27], while the FDA indicates that "for adjunctive treatment, studies should include patients with partial responses to other antidepressant therapies" [26].

Consistent with the most commonly used criteria, we recommend that TRD should be defined after a minimum of two failed treatments with <25% of improvement with adequate dosing and duration (strong consensus, 96%), while PRD can be defined even after a single treatment (improvement 25 to <50%) with adequate dosing and duration (moderate consensus, 76%).

Current or past episodes
The preferred definition of TRD for clinical trials includes a current failure and a past failure, i.e., subjects are currently receiving an antidepressant and they are still depressed according to current clinician's assessment, and they were also treated with another antidepressant in the past and showed no response based on retrospective assessment. However, it is unclear whether both of these treatments should apply to the same (current) episode or to clearly distinct episodes. Indeed, although the EMA definition of MDD [27] emphasizes the current episode for the characterisation of the disease, it does not clarify whether the two failures should both be during the current episode. Of course, it is difficult to retrospectively defining the response to an antidepressant, especially if the current episode is of long duration (years) and the previous treatment was closer to the onset of the episode.

We recommend (with weak consensus, 51%) that the definition of TRD should include two treatment failures both within the current episode, and the definition of PRD should include partial response to at least one treatment within the current episode; moreover, for long current episodes, only treatment failures within the last two years should be considered. Of note, the "minority" position was split across a continuum of different opinions ranging from those who wanted to maintain that both antidepressants should be in the same episode but within a shorter period of time (shorter than two years) to those who wanted to consider two antidepressants "life-time"; thus, our recommendation sits in the middle of these extremes.

Prospective or retrospective assessment
A related issue, particularly important for regulatory clinical trials, is whether one treatment failure should be 'prospective within the trial', i.e., the trial starts with an established medication for MDD at an adequate dose, and then the person is offered a new intervention only if the medication fails to improve or eliminate MDD symptoms. We acknowledge that this, while ideal, would lead to operational execution challenges within a trial, with increased complexity and burden for the sites and study participants [22].

We recommend (with moderate consensus, 75%) that it is possible to assess ineffective past/current antidepressant treatment attempts, but only if properly documented, that is, based not only on subjective recollection or standardised instruments to assess psychiatric history and previous treatments (see below), but also on clinical documentation, such as pharmacy, hospital, or other health records. This documentation can also be used to confirm some degree of adherence to the failed treatments, and to screen people with depressive symptoms for previous episodes of mania, hypomania, or sub-threshold bipolarity, since these individuals should be excluded (see below).

Adherence to treatment is a well-recognized critical issue in both clinical and research settings. Rates of adherence vary across the literature and generally are limited by different and/or restricted time periods [28]. However, people with MDD have typically high reported rates of non-adherence (up to more than 50%) [29, 30]. It is therefore vital to properly confirm the individual's adherence in order to define non-response: many cases of TRD may not be true TRD, but, instead, represent partial or full non-adherence. However, the assessment of treatment adherence can be difficult and is often not addressed in everyday settings, or in most current studies in TRD. The most reliable method to assess adherence is to perform a blood test to measure the concentration of the medication in the plasma (which would also allow the recognition of fast and slow metabolisers), even though it may increase the participants' burden during the trial. Although this suggestion was clearly supported by the experts, systematic use of plasma level monitoring in TRD/PRD definitions is not current practice. It is important to note that the FDA does not accept data analyses that exclude individuals not compliant to the previous treatment based on a blood dosing of the medicine. Further discussion on adherence is presented in the Appendix. However, we recommend assessing the usefulness of different methods to measure a person's adherence for a potential (and desirable) future implementation (see Future directions).
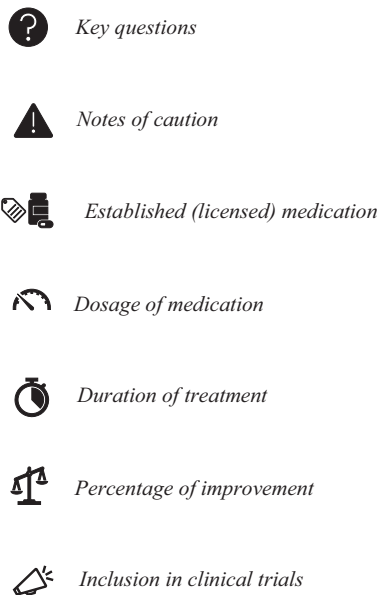
TYPE OF MEDICATIONS
Different classes and mechanisms of action
The EMA [27] mentions that the two treatment failures could be with medications of "same or different class". Other guidelines,
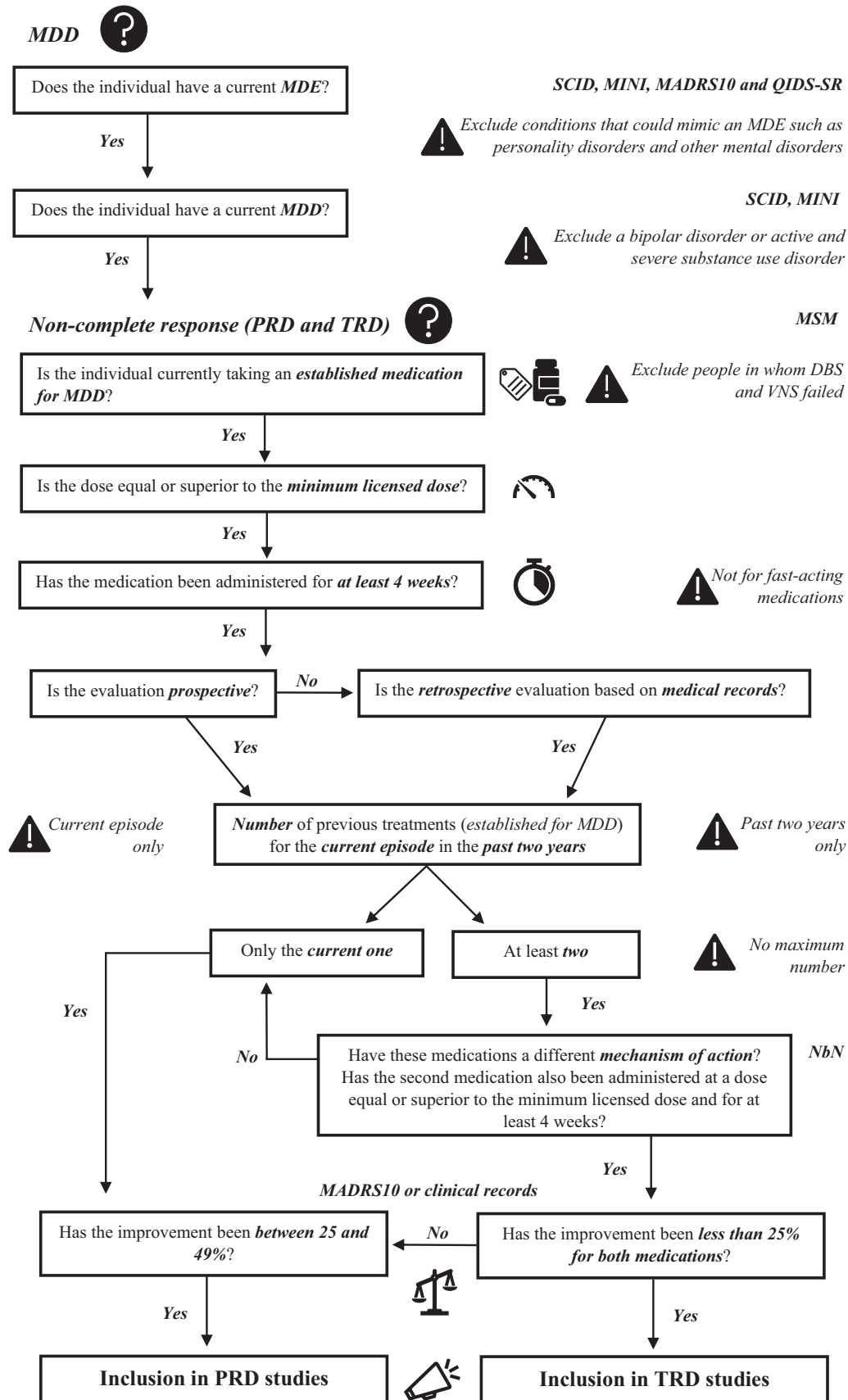
Fig. 2 Algorithm for inclusion in TRD/PRD studies. Recommended algorithm for inclusion of participants in TRD and PRD studies based on the consensus recommendations discussed throughout the manuscript and listed in Table 1.

DBS= deep brain stimulation; MADRS10= Montgomery-Asberg Depression Rating Scale; MDD= major depressive disorder; MDE= major depressive episode; MINI= Mini-International Neuropsychiatric Interview; MSM= Maudsley Staging Model; NbN= Neuroscience-based Nomenclature; PRD= partially responsive depression; QIDS-SR= Quick Inventory of Depressive Symptomatology Self-Report; SCID= Structured Clinical Interview for DSM; TRD= treatment-resistant depression; VNS= vagus nerve stimulation.

Key questions

Notes of caution

Established (licensed) medication

Dosage of medication

Duration of treatment

Percentage of improvement

Inclusion in clinical trials

The first key question to answer is whether the individual has an MDE in the context of an MDD, using validated clinical scales (MADRS10 and QIDS-SR) and diagnostic instruments (SCID and MINI). Important is to exclude: personality disorders, bipolar disorder or active and severe substance use disorder; and people in whom DBS and VNS failed to improve/eliminate MDD symptoms. A second key question is whether the individual has a non-complete response (TRD or PRD), assessed ideally through the MSM. The individual needs to be currently taking an established (licensed) medication for MDD at a dosage equal or superior to the minimum licensed dose and administered for at least 4 weeks (this does not apply to fast-acting medications). The evaluation should be prospective, or, when retrospective, based on health records. There is no maximum number of previous antidepressants to consider. For inclusion in TRD studies (on the right), there should be at least two previous treatments (both established for MDD), both in the current episode and in the past two years, with an improvement less than 25% for both medications (based on MADRS10 or clinical records). These medications must have a different mechanism of action (according to the NbN) and the second medication must have also been administered at a dose equal or superior to the minimum licensed dose and for at least 4 weeks. For inclusion in PRD studies (on the left), is sufficient one previous treatment (established for MDD), for the current episode and in the past two years, with an improvement between 25 and 49% (based on MADRS10 or clinical records).

Fig. 2 (continue)

including the FDA guidance [26], do not make this distinction. Interestingly, the concept of 'different mechanisms' may overlap with 'different classes', although the pharmacological overlap between classes and mechanism of action is not absolute; the Neuroscience-based Nomenclature (NbN) for psychotropic agents has greatly contributed to clarify this issue [31]. Moreover, there are some geographically-relevant label limitations to the usage of some medications (see also Appendix).

We recommend (with moderate consensus, 75%) that, to define TRD, the two antidepressant treatment failures should be with two established (licensed) medications for MDD of different mechanisms of action. The two treatments could be separated by a drug free period (and there is failure to both), prescribed sequentially (switching, with failure to both), or one is prescribed as augmentation to the other (because of the first drug failing, and with the augmentation also failing).

Moreover, we also recommend (with moderate consensus, 78%) that a failed course of psychotherapy should not be included as one of the previous treatments required for the definition of TRD/PRD, but this information should be reported for staging (see below). Notably, this recommendation does not have implications for the role of psychotherapies in the treatment of TRD, and, similarly to other TRD models proposed [11], we support the use of psychotherapies as a treatment strategy for TRD in clinical practice.

Dosage and duration of antidepressants
Most studies and meta-analyses have found no benefit for antidepressant dose escalation versus staying on the minimum licensed dose, with an increased risk of side effects and discontinuation [32, 33], even if there is also evidence that higher doses may have superior efficacy [34, 35] and that higher starting doses may be associated with higher response rates [36]. In terms of 'minimum duration', some studies have shown that, while two weeks can be enough to observe an initial response, a stable response/remission is usually detectable after four weeks [37, 38].

However, a complete remission may not be detectable after four weeks. For example, approximately half of remitters on citalopram in level 1 of the STAR*D trial remitted between week 6 and week 14 [39]. Of note, this is different for fast-acting medications, such as esketamine [40], but these drugs are indicated for individuals that already have TRD.

We recommend (with moderate consensus, 74%) that the criteria of 'adequate dose and duration' is the minimal effective dosage, that is, the minimal approved dosage, administered for at least four weeks.

Interestingly, in regulatory clinical trials, participants' discontinuation of treatments may be considered as an endpoint equivalent to a failure [41]. However, we recommend (with strong consensus, 96%) that discontinuation of treatment before the completion of the fourth week, without clear evidence of lack of response, should not be considered as a treatment failure for the purpose of establishing TRD/PRD, especially because of the difficulty in distinguishing retrospectively between non-response and intolerance.

EXCLUSION FROM TRD/PRD STUDIES
Many people with lived experience (PWLE) who fulfil the proposed criteria for TRD may have tried many ineffective antidepressants (multi-drug resistance) or other (licensed or unlicensed) pharmacological augmentation interventions (see also Appendix). We discussed whether these individuals should be considered to have TRD which is 'too severe' for inclusion in clinical trials.

We recommend (with moderate consensus, 93%) that multiple-drug resistant individuals, and individuals in whom augmentation strategies failed to improve/eliminate MDD symptoms should not be excluded from TRD/PRD studies, provided the other criteria are met. We also recommend (with strong consensus, 100%) that a previous structured psychotherapy failing to improve MDD symptoms is not an exclusion criterion. Rather, all this information should be recorded for staging and for potential subgroup analyses.

Deep brain stimulation (DBS) and vagus nerve stimulation (VNS) are continuous treatments in which a neurostimulator (usually implanted in the chest wall) is connected to intracerebral electrodes or the left vagus nerve [42, 43]. Because of the invasiveness of these intervention, people selected for these treatments have tried an unusually large number of ineffective treatments, and thus are different from other people living with TRD [44, 45].

We recommend (with moderate consensus, 62%) that individuals with MDD in whom DBS and VNS failed to improve/eliminate MDD symptoms should be excluded from TRD/PRD clinical studies. However, we also recommend (with moderate consensus, 88%) that individuals with MDD in whom other non-continuous/non-invasive brain stimulation interventions, such as electroconvulsive therapy (ECT) or transcranial magnetic stimulation (TMS), failed to improve/eliminate MDD symptoms should not be excluded.

CLINICAL PRESENTATION
TRD and PRD symptoms
People living with MDD have a range of symptoms specifiers (for example, with melancholic features, with atypical features, with psychotic features), as well as the presence of comorbid diagnoses (for example, comorbid anxiety, bipolar depression). Potentially, some of these symptoms may be more difficult to treat compared with others, such as MDD with anxious distress [46–49].

We recommend (with strong consensus, 95%) that all specifiers of depression (melancholic, atypical, anxious, psychotic, mixed) should be considered within the TRD/PRD definition, except for bipolar depression, which should be excluded as this is part of bipolar disorder.

Comorbidities
Individuals with a diagnosis of personality disorder (especially borderline personality disorder) frequently meet criteria for MDD, but antidepressants are unlikely to be effective and hence may mimic a non-response. Active substance users also present an increased risk of pharmacological interactions (both pharmacokinetic and pharmacodynamic), side effects, and mood symptoms due to the effects of the substance or of the withdrawal which may appear as a 'phenocopy' of MDD. However, it is important to highlight that these conditions are so frequently comorbid with depression that they cannot be routinely excluded, otherwise the proposed TRD/PRD definitions would not be generalisable to individuals seen in everyday clinics.

The FDA guidance also captures this tension between homogeneity and generalisability, stating that "investigators should seek demographically broad populations and avoid unnecessary restriction of study populations (e.g., by excluding patients with concomitant illness)" [26]. Indeed, the FDA does not explicitly take position on the inclusion or exclusion of personality disorders, and it explicitly encourages to consider people with a history of substance abuse, "although such inclusions should be weighed against concerns about diagnostic and medication effect confounders", and further states that "patients whose substance use disorder is not at least in partial remission will likely be excluded from antidepressant trials depending on the level of particular confounding concerns". On the other hand, the EMA document only broadly indicates that MDD occurring comorbid with other psychiatric disorders is not in the remit of the guideline [27].

We recommend (with moderate consensus, 79%) that comorbid personality disorders or other mental disorders should be excluded from TRD/PRD studies only when their onset is properly documented as independent and antecedent to the MDD diagnosis. Moreover, in accordance with the FDA guidance [26], we recommend (with moderate consensus, 83%) that individuals with a severe substance use disorder not currently in remission should be excluded from TRD/PRD studies, independently from the onset; in contrast, individuals with comorbid substance use disorder that is active and mild/moderate should be excluded from TRD/PRD studies only when the onset is properly documented as independent and antecedent to the MDD diagnosis.

Also, somatic comorbidities should be systematically recorded, but not excluded a priori; this includes conditions such as inflammatory, neuroendocrine, and metabolic diseases, which can influence the response to treatments [50–52].

DIAGNOSTIC TOOLS AND MEASURES OF OUTCOME
We also established a consensus on the best psychometric tools to measure antidepressant response (and thus TRD/PRD status and staging) both retrospectively, to improve the way we diagnose TRD (and PRD) before entering a trial, and prospectively, as tools to use in regulatory trials. A detailed summary of the different psychometric instruments is beyond this core document and is instead presented in the Appendix.

Historical assessment of TRD/PRD status
At a minimum, assessment to define TRD should include a structured clinical interview for the diagnosis of MDD, such as the Structured Clinical Interview for DSM (SCID) [53, 54] and the Mini-International Neuropsychiatric Interview (MINI) [55], together with a scale to assess the patient's antidepressant history, such as the Massachusetts General Hospital Antidepressant Treatment Response Questionnaire (ATRQ) [56], and the Antidepressant Treatment History Form (ATHF) [57]. A more structured method is to use staging models [6, 58], such as the Thase and Rush method [59], the Massachusetts General Hospital Staging model (MGH-s) [22], and the more recent Maudsley Staging Model (MSM) [60], which also allows

the assessment of previous antidepressant treatment failure, using the Maudsley Treatment Inventory (MTI) [10].

Among the experts, there was a substantial agreement that staging models are the preferred instrument to define TRD/PRD and assess the level of treatment-resistance, even if many experts recognize the validity of the ATRQ, often accepted in regulatory clinical trials.

We recommend (with moderate consensus, 69%) the Maudsley Staging Model as the preferred instrument. However, both the Thase and Rush and the MGH-s models are valid alternatives.

Assessment of depressive symptoms and response to antidepressant treatment

The presence and severity of depression are established through clinician-administered scales, such as the Hamilton Depression Rating Scale (HAM-D 17, 21, 24 items) [61], the Montgomery-Åsberg Depression Rating Scale (MADRS10) [62], and the Quick Inventory of Depressive Symptomatology (QIDS) Clinician Rating (QIDS-C) [63]; or self-reported instruments, such as the Beck Depression Inventory (BDI) [64], the Patient Health Questionnaire-9 Item (PHQ-9) [65], and the QIDS Self-Report (QIDS-SR) [63]. When possible, these scales are administered before and after a specific treatment has started, but they can also be used as a single measure to determine the current severity of the depression. Although TRD/PRD definitions centre on 'response', the ideal treatment goal is remission, considered as the absence of a relevant MDD symptomatology as indicated by a score of ≤7 at the HAM-D17 [66] or ≤10 at the MADRS10 [67, 68].

Of note are also the Clinical Global Impression (CGI) scales, measuring psychiatric global status [69], and the Sheehan Disability Scale (SDS), assessing functional impairment [70], which are important since even responders may continue to have significant residual symptoms and functional impairment. Indeed, people with lived experience of depression often define treatment success with an emphasis on broadly-based functional outcomes rather than remission of individual symptoms, and their perspective should be included through patient-reported outcomes (PROs) such as quality of life, wellbeing, impact of symptoms, and social and occupational functioning [71].

We recommend (with moderate consensus) that the clinician administered MADRS10 is the preferred outcome instrument to assess treatment response (and remission) (85%), together with patient-reported QIDS-SR (81%). We additionally recommend (with moderate consensus, 92%) that criteria for remission, response, and partial response should not be relaxed in regulatory clinical trials for TRD/PRD and shorter versions of the traditional scales, such as the HAM-D6 and the MADRS6, should not be preferred to full scales.

The MADRS10 and QIDS-SR were chosen based solely on experts' feedback, and not a priori selected (see also Appendix). The MADRS10 was originally developed with the precise aim to have a scale that was more sensitive to change compared with the HAM-D17, a crucial feature for clinical trials [72]. In addition, unlike the HAM-D [73], it predominately focuses on the core symptoms of depression such as sadness, tension, lassitude, pessimistic thoughts, and suicidal thoughts. This results in a higher internal consistency and a greater accuracy for the MADRS compared with HAM-D [74]. However, this is also a potential limitation of the MADRS, as it may not capture the clinical complexity of the PWLE. Most notably, individuals with MDD with non-melancholic or atypical features may present with a range of symptoms that are not captured by the MADRS. For such reason, there is a risk that MDD individuals continuing to report depressive symptoms such as irritability and anxiety may be classified as responders. The same issue may arise with a 'short' scale such as the QIDS-SR, which again carries the risk of omitting the measure of MDD symptoms that are not captured by standard diagnostic criteria of MDD, in contrast to far more granular self-report measures of depressive symptoms, such as the Inventory of Depressive

Symptomatology (IDS)-SR [75], and the Symptoms of Depression Questionnaire (SDQ) [76]. Thus, the assessment of changes in MDD severity (upon which the TRD/PRD definitions rely) should include the entirety of symptoms that an individual displays, including different (non-melancholic) MDD specifiers and symptoms not included in standard classification systems. When possible, the broadest assessment possible should be conducted, using broad-spectrum scales or checklists assessing a wide range of MDD related symptoms, such as the HAM-D28. Also, specific instruments assessing specific symptoms, such as anhedonia, motor retardation, or anxiety (for example, the Hamilton Anxiety Rating Scale (HAM-A) [77]), may be useful. Sub-analyses on specific clusters of symptoms (such as suicidal, atypical, or psychotic symptoms) could be used to generate hypotheses for future clinical trials, for medications targeted to specific symptoms or to symptoms-based subgroups of individuals. Another issue is that traditional scales address depressive symptoms in the emotional and physical clusters, with a lower attention on the cognitive ones, which may frequently remain as residual symptoms indicating poor response and thus an increased risk for relapse. Interestingly, there is evidence that a mismatch between scores in clinician-administered and self-reported scales is a poor prognostic sign [78], supporting the position endorsed in this report to have at least one score of both types of scales included, until new and more tailored instruments will be developed (see also next section and Appendix).

FUTURE DIRECTIONS

This section makes recommendations that are not immediately applicable to clinical trials and highlights research gaps and new opportunities identified throughout the consultative process. Future trial sponsors may wish to incorporate one or more of these areas to expand the knowledge base. A more detailed discussion of these points is presented in the Appendix; here, we want to emphasise the main recommendations:

- Future research should recognize and target different clinical phenotypes of TRD (and PRD) underpinned by a specific biological mechanism (strong, 100%).
- For future research, diagnostic and history-taking instruments should be implemented in clinical cohorts and electronic health records, to allow a reliable, comprehensive, and multidimensional evaluation of the PWLE (strong, 100%).
- Currently, no biomarker has been validated in clinical practice or in clinical trials to identify subjects with TRD (and PRD), or to further stratify them; however, collection of biological samples for subsequent subgroup or stratified analyses is recommended (moderate, 91%).
- Preferences, perspectives, and reported outcomes of PWLE should be included in future TRD (and PRD) diagnostic tools and outcome measures (strong, 100%).
- The usefulness of adherence assessment using blood levels or other methods (also in a run-in period) should be assessed through research, before deciding whether it should be implemented in future clinical trials (moderate, 92%).

LIMITATIONS AND CONCLUSIONS

While we discuss many points of debate in the Appendix, here we want to acknowledge the main limitations of this report.

First, the recommendations are based on experts' and stakeholders' clinical and professional judgement, supported by research and clinical evidence, but without any hard and objective validation. Second, our recommendations try to strike a balance between being 'too broad' and being 'too narrow', with the aim to identify 'clinically homogenous' samples, based on a pragmatic

and non-aetiologically-based approach; still, we cannot exclude that significant phenotypic or biological heterogeneity is present within the defined TRD/PRD samples, and future research will allow to define this inherent variability, if any, in clinical features, biomarkers, and clinical response. Third, the opposite risk is also present: that we define groups so narrowly that the findings are not generalisable to the larger population of PWLE in clinical settings, and do not translate into better care; however, maintaining narrow definitions may help avoiding overmedicalisation and may prevent the lowering of the threshold above which treatments which are specific for TRD are offered to patients, as a recent commentary on this topic highlighted [18].

In terms of theoretical limitations, some of our contributors have argued that the proposed TRD/PRD definitions are unhelpful from a clinical and a conceptual perspective, as they arbitrarily apply thresholds on a continuum, and are influenced by the different healthcare systems. Moreover, TRD and PRD concepts are based upon a conceptualisation of depression as an episodic illness with good inter-episode recovery, while many individuals with MDD have a chronic illness with waxing and waning course; this might create difficulties when trying to accurately identify the beginning and end of the current/previous episodes, in order to define response to the antidepressants. Finally, we recognise that, according to our recommendations, individuals with MDD with an ineffective (<25%) single antidepressant trial are not defined as either TRD or PRD; as in standard clinical practice, these subjects should be prescribed a second antidepressant before a decision can be made on their TRD status.

Notwithstanding these limitations (and the points of uncertainties further discussed in the Appendix), this report offers clear and consistent definitions of TRD/PRD for regulatory clinical trials and for clinical and biological studies more broadly, agreed among a large group of experts including clinicians, academics, industry, regulatory agencies, and PWLE. Our ultimate ambition is to advance tailored treatments and a truly 'precision medicine' approach for MDD, which in turn will finally help to deliver better care for people suffering from this severely challenging illness, which remains too often ineffectively treated.

DISCLAIMER
This document reflects the majority views of the experts. Authorship reflects having been part of the process and accepting the consensus statements without necessarily personally endorsing every single recommendation (see level of consensus). Also, the views expressed in this article are the personal views of the author(s) and may not be understood or quoted as being made on behalf of or reflecting the position of the regulatory agency/agencies or organisations with which the author(s) is/are employed/affiliated. This publication reflects only the authors' view, and the JU is not responsible for any use that may be made of the information it contains.

REFERENCES
1. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet. 2018;391:1357–66.
2. Otte C, Gold SM, Penninx BW, Pariante CM, Etkin A, Fava M, et al. Nature disease primer: Major Depressive Disorder. Nature Publishing Group. 2016.
3. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. Am J Psychiatry. 2006;163:1905–17.
4. McIntyre RS, Filteau MJ, Martin L, Patry S, Carvalho A, Cha DS, et al. Treatment-resistant depression: Definitions, review of the evidence, and algorithmic approach. J Affect Disord. 2014;156:1–7.
5. Gaynes BN, Lux L, Gartlehner G, Asher G, Forman-Hoffman V, Green J, et al. Defining treatment-resistant depression. Depression Anxiety. 2019;37:134–45.
6. Salloum NC, Papakostas GI. Staging treatment intensity and defining resistant depression: Historical overview and future directions. Journal of Clinical Psychiatry. 2019, 80.
7. McAllister-Williams RH, Arango C, Blier P, Demyttenaere K, Falkai P, Gorwood P, et al. The identification, assessment and management of difficult-to-treat depression: An international consensus statement. J Affect Disord. 2020;267:264–82.
8. Sheehan DV, Harnett-Sheehan K, Spann ME, Thompson HF, Prakash A. Assessing remission in major depressive disorder and generalized anxiety disorder clinical trials with the discan metric of the Sheehan disability scale. Int Clin Psychopharmacol. 2011;26:75–83. Mar.
9. McAllister-Williams RH, Christmas DMB, Cleare AJ, Currie A, Gledhill J, Insole L, et al. Multiple-therapy-resistant major depressive disorder: A clinically important concept. Br J Psychiatry. 2018;212:274–8.
10. Fekadu A, Donocik JG, Cleare AJ. Standardisation framework for the Maudsley staging method for treatment resistance in depression. BMC Psychiatry. 2018, 18.
11. Conway CR, George MS, Sackeim HA. Toward an evidence-based, operational definition of treatment-resistant depression: When Enough is enough. JAMA Psychiatry Am Med Assoc. 2017;74:9–10.
12. Brown S, Rittenbach K, Cheung S, McKean G, MacMaster FP, Clement F. Current and Common Definitions of Treatment-Resistant Depression: Findings from a Systematic Review and Qualitative Interviews. Can J Psychiatry SAGE Publ Inc. 2019;64:380–7.
13. Howes OD, Thase ME, Pillinger T. Treatment resistance in psychiatry: state of the art and new directions. Mol Psychiatry 2021. 2021;13:1–15. Jul.
14. Taylor RW, Marwood L, Oprea E, Deangel V, Mather S, Valentini B, et al. Pharmacological Augmentation in Unipolar Depression: A Guide to the Guidelines. International Journal of Neuropsychopharmacology. 2021.
15. Strawbridge R, Carter B, Marwood L, Bandelow B, Tsapekos D, Nikolova VL, et al. Augmentation therapies for treatment-resistant depression: Systematic review and meta-analysis. Br J Psychiatry. 2019;214:308–308.
16. McAllister-Williams RH, Arango C, Blier P, Demyttenaere K, Falkai P, Gorwood P, et al. Reconceptualising treatment-resistant depression as difficult-to-treat depression. The Lancet. Psychiatry 2021;8:14–15.
17. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med. 2010;7:e1000217.
18. Cosgrove L, Naudet F, Högberg G, Shaughnessy AF, Cristea IA. Reconceptualising treatment-resistant depression as difficult-to-treat depression. The Lancet. Psychiatry 2021;8:11–3.
19. Humphrey-Murto S, de Wit M. The Delphi method—more research please. J Clin Epidemiol Elsevier USA. 2019;106:136–9.
20. Grant S, Booth M, Khodyakov D. Lack of preregistered analysis plans allows unacceptable data mining for and selective reporting of consensus in Delphi studies. J Clin Epidemiol. 2018;99:96–105. Jul 1.
21. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol. 2014;67:401–9. Apr.
22. Fava M. Diagnosis and definition of treatment-resistant depression. Biol Psychiatry. 2003;53:649–59.
23. Hirschfeld RMA, Montgomery SA, Aguglia E, Amore M, Delgado PL, Gastpar M, et al. Partial response and nonresponse to antidepressant therapy: Current approaches and treatment options. J Clin Psychiatry. 2002;63:826–37.
24. Kennedy SH, Lam RW, McIntyre RS, Tourjman SV, Bhat V, Blier P, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: Section 3. Pharmacological Treatments. Canadian Journal of Psychiatry. 2016.
25. Jackson WC, Papakostas GI, Rafeyan R, Trivedi MH. Recognizing Inadequate Response in Patients With Major Depressive Disorder. The Journal of clinical psychiatry. 2020, 81.
26. Food and Drug Administration. Major Depressive Disorder: Developing Drugs for Treatment, Guidance for Industry', U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Revision 1. 2018.
27. European Medicines Agency. Guideline on clinical investigation of medicinal products in the treatment of depression. EMA/CHMP/185423/2010 Rev 2. 2013.
28. Keyloun KR, Hansen RN, Hepp Z, Gillard P, Thase ME, Devine EB. Adherence and Persistence Across Antidepressant Therapeutic Classes: A Retrospective Claims Analysis Among Insured US Patients with Major Depressive Disorder (MDD). CNS Drugs. 2017;31:421–32.
29. ten Doesschate MC, Bockting CLH, Schene AH. Adherence to continuation and maintenance antidepressant use in recurrent depression. J Affect Disord. 2009;115:167–70.
30. Ho SC, Chong HY, Chaiyakunapruk N, Tangiisuran B, Jacob SA. Clinical and economic impact of non-adherence to antidepressants in major depressive disorder: A systematic review. J Affect Disord. 2016;193:1–10.

1296

31. Frazer A, Blier P. A Neuroscience-based Nomenclature (NbN) for psychotropic agents. Int J Neuropsychopharmacol. 2016;19:pyw066.

32. Cleare A, Pariante CM, Young AH, Anderson IM, Christmas D, Cowen PJ, et al. Evidence-based guidelines for treating depressive disorders with antidepressants: A revision of the 2008 British Association for Psychopharmacology guidelines. J Psychopharmacol 2015. 2015;29:459–525.

33. Furukawa TA, Salanti G, Cowen PJ, Leucht S, Cipriani A. No benefit from flexible titration above minimum licensed dose in prescribing antidepressants for major depression: systematic review. Acta Psychiatr Scandinavica. 2020;141: 401–9.

34. Hieronymus F, Nilsson S, Eriksson E. A mega-Analysis of fixed-dose trials reveals dose-dependency and a rapid onset of action for the antidepressant effect of three selective serotonin reuptake inhibitors. Transl Psychiatry. 2016;6:e834–e834.

35. Jakubovski E, Varigonda AL, Freemantle N, Taylor MJ, Bloch MH. Systematic review and meta-analysis: Dose-response relationship of selective serotonin reuptake inhibitors in major depressive disorder. Am J Psychiatry. 2016;173:174–83.

36. Papakostas GI, Charles D, Fava M. Are typical starting doses of the selective serotonin reuptake inhibitors sub-optimal? A meta-analysis of randomized, double-blind, placebo-controlled, dose-finding studies in major depressive disorder. World J Biol Psychiatry. 2010;11:300–7.

37. Szegedi A, Jansen WT, Van Willigenburg APP, Van Der Meulen E, Stassen HH, Thase ME. Early improvement in the first 2 weeks as a predictor of treatment outcome in patients with major depressive disorder: A meta-analysis including 6562 patients. J Clin Psychiatry. 2009;70:344–53.

38. Kelley ME, Dunlop BW, Nemeroff CB, Lori A, Carrillo-Roa T, Binder EB, et al. Response rate profiles for major depressive disorder: Characterizing early response and longitudinal nonresponse. Depression Anxiety. 2018;35:992–1000.

39. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, et al. Evaluation of Outcomes With Citalopram for Depression Using Measurement-Based Care in STAR*D. Implic Clin Pract. 2006;163:28–40. Jan 1.

40. McIntyre RS, Rosenblat JD, Nemeroff CB, Sanacora G, Murrough JW, Berk M, et al. Synthesizing the Evidence for Ketamine and Esketamine in Treatment-Resistant. Depression: Int Expert Opin Available Evid Implement. 2021;178:383–99. Mar 17.

41. Pradier MF, McCoy TH, Hughes M, Perlis RH, Doshi-Velez F. Predicting treatment dropout after antidepressant initiation. Transl Psychiatry. 2020;10:60.

42. Bauer M, Severus E, Köhler S, Whybrow PC, Angst J, Möller HJ. World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders. part 2: Maintenance treatment of major depressive disorder-update 2015. World Journal of Biological Psychiatry. 2015.

43. Dandekar MP, Fenoy AJ, Carvalho AF, Soares JC, Quevedo J. Deep brain stimulation for treatment-resistant depression: an integrative review of preclinical and clinical findings and translational implications. Mol Psychiatry. 2018;23:5. 2018 Feb 27;23:1094–112.

44. Holtzheimer PE, Husain MM, Lisanby SH, Taylor SF, Whitworth LA, McClintock S, et al. Subcallosal cingulate deep brain stimulation for treatment-resistant depression: a multisite, randomised, sham-controlled trial. The Lancet. Psychiatry 2017;4:839–49.

45. Rush AJ, Siefert SE. Clinical issues in considering vagus nerve stimulation for treatment-resistant depression. Exp Neurol. 2009;219:36–43.

46. Souery D, Oswald P, Massat I, Bailer U, Bollen J, Demyttenaere K, et al. Clinical factors associated with treatment resistance in major depressive disorder: Results from a European multicenter study. J Clin Psychiatry. 2007;68:1062–70.

47. Rush AJ, Wisniewski SR, Warden D, Luther JF, Davis LL, Fava M, et al. Selecting among second-step antidepressant medication monotherapies: Predictive value of clinical, demographic, or first-step treatment features. Arch Gen Psychiatry. 2008;65:870–80.

48. Bartova L, Dold M, Kautzky A, Fabbri C, Spies M, Serretti A, et al. Results of the European Group for the Study of Resistant Depression (GSRD) — basis for further research and clinical practice. World J Biol Psychiatry. 2019;20:427–48.

49. Braund TA, Palmer DM, Williams LM, Harris AWF. Characterising anxiety in major depressive disorder and its use in predicting antidepressant treatment outcome: An iSPOT-D report. Aust NZ J Psychiatry. 2019;53:782–93.

50. Juruena MF, Pariante CM, Papadopoulos AS, Poon L, Lightman S, Cleare AJ. The role of mineralocorticoid receptor function in treatment-resistant depression. J Psychopharmacol. 2013;27:1169–79.

51. Strawbridge R, Arnone D, Danese A, Papadopoulos A, Herane Vives A, Cleare AJ. Inflammation and clinical response to treatment in depression: A meta-analysis. Eur Neuropsychopharmacol. 2015;25:1532–43.

52. Pan LA, Martin P, Zimmer T, Segreti AM, Kassiff S, McKain BW, et al. Neurometabolic disorders: Potentially treatable abnormalities in patients with treatment-refractory depression and suicidal behavior. Am J Psychiatry. 2017;174:42–50.

53. First, MB, Spitzer, RL, Gibbon M, and Williams JBW. Structured Clinical Interview for DSM-IV-TR Axis I Disorders. New York State Psychiatric Institute. 2002.

54. First MB, Williams JBW, Karg RS, Spitzer RL. Structured clinical interview for DSM-5 research version. American Psychiatric Association, Washington DC. 2015.

55. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. In: Journal of Clinical Psychiatry. 1998.

56. Chandler GM, Iosifescu DV, Pollack MH, Targum SD, Fava M. Validation of the massachusetts general hospital Antidepressant Treatment History Questionnaire (ATRQ). CNS Neurosci Therapeutics. 2010;16:322–5.

57. Sackeim HA. The definition and meaning of treatment-resistant depression. J Clin Psychiatry. 2001;62:10–7. Suppl 16.

58. Ruhé HG, Van Rooijen G, Spijker J. Peeters FPML, Schene AH. Staging methods for treatment resistant depression. A systematic review. J Affect Disord. 2012;137: 35–45.

59. Thase ME, Rush AJ. When at first you don't succeed: Sequential strategies for antidepressant nonresponders. In: Journal of Clinical Psychiatry. 1997.

60. Fekadu A, Wooderson S, Donaldson C, Markopoulou K, Masterson B, Poon L, et al. A multidimensional tool to quantify treatment resistance in depression: The Maudsley staging method. J Clin Psychiatry. 2009;70:177–84.

61. Hamilton M. A rating scale for depression. J Neurol, Neurosurg, psychiatry. 1960;23:56–62.

62. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134:382–9.

63. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. Biol Psychiatry. 2003;54:573–83.

64. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An Inventory for Measuring Depression. Arch Gen Psychiatry. 1961;4:561–71.

65. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med. 2001;16:606–13.

66. Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recovery, relapse, and recurrence. Arch Gen Psychiatry. 1991;48:851–5.

67. Hawley CJ, Gale TM, Sivakumaran T. Defining remission by cut off score on the MADRS: Selecting the optimal value. J Affect Disord. 2002;72:177–84.

68. Zimmerman M, Posternak MA, Chelminski I. Defining remission on the montgomery-asberg depression rating scale. J Clin Psychiatry. 2004;65:163–8.

69. Guy W. Rating Clinician-rated | Clinical Global Impression (CGI). ECDEU Assessment Manual for Psychopharmacology. 1976.

70. Leon AC, Olfson M, Portera L, Farber L, Sheehan DV. Assess Psychiatr Impairment Prim Care Sheehan Disabil Scale. 1997;27:93–105. Jul 1.

71. Deshpande P, Bl Sudeepthi, Rajan S, Abdul Nazir C. Patient-reported outcomes: A new era in clinical research. Perspect Clin Res. 2011;2:137–44.

72. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134:382–9.

73. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: Has the gold standard become a lead weight? Am J Psychiatry Am Psychiatr Publ. 2004;161:2163–77.

74. Uher R, Farmer A, Maier W, Rietschel M, Hauser J, Marusic A, et al. Measuring depression: Comparison and integration of three scales in the GENDEP study. Psychological Med. 2008;38:289–300. Feb.

75. Trivedi MH, Rush AJ, Ibrahim HM, Carmody TJ, Biggs MM, Suppes T, et al. The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. Psychological Med. 2004;34:73–82. Jan.

76. Pedrelli P, Blais MA, Alpert JE, Shelton RC, Walker RSW, Fava M. Reliability and validity of the Symptoms of Depression Questionnaire (SDQ). CNS Spectr. 2014;19:535–46. Dec 23.

77. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol. 1959;32:50–5.

78. Dunlop BW, Li T, Kornstein SG, Friedman ES, Rothschild AJ, Pedersen R, et al. Concordance between clinician and patient ratings as predictors of response, remission, and recurrence in major depressive disorder. J Psychiatr Res. 2011;45: 96–103.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

COMPETING INTERESTS

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-021-01381-x.

Correspondence and requests for materials should be addressed to Luca Sforzini.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]King's College London, Institute of Psychiatry, Psychology and Neuroscience, Department of Psychological Medicine, London, UK. [2]Neuroscience and Psychiatry Unit, Division of Neuroscience and Experimental Psychology, School of Biological Sciences, Faculty of Biological, Medical and Human Sciences, The University of Manchester and Manchester Academic Health Sciences Centre, Manchester, UK. [3]Department of General and Academic Psychiatry, Centre Hospitalier Charles Perrens, Laboratory of Nutrition and Integrative Neurobiology (UMR INRAE 1286), University of Bordeaux, Bordeaux, France. [4]Department of Psychiatry and Psychotherapy, University of Münster, Münster, Germany. [5]Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany. [6]Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, Medical Faculty, Technische Universität Dresden, Dresden, Germany. [7]Department of Psychiatry, Melbourne Medical School and The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia. [8]Mood Disorders Research Unit, The Royal's Institute of Mental Health Research, Ottawa, ON, Canada. [9]Department of Psychiatry, University of Ottawa, Ottawa, ON, Canada. [10]Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada. [11]National Institute for Health Research Mental Health Biomedical Research Centre, South London and Maudsley NHS Foundation Trust and King's College London, London, UK. [12]Medical Sciences Division, Department of Psychiatry, University of Oxford, Oxford, UK. [13]APC Microbiome Ireland, Cork, Ireland; Department of Psychiatry and Neurobehavioral Sciences, University College Cork, Cork, Ireland. [14]Department of Molecular Medicine, Division of Psychiatry, University of Siena, Siena, Italy. [15]Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK. [16]Depression Research Center of the German Depression Foundation and Department of Psychiatry, Psychosomatics and Psychotherapy, Goethe University, Frankfurt, Germany. [17]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA. [18]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, USA. [19]Université Paris Est Creteil (UPEC), AP-HP, Hôpitaux Universitaires Henri Mondor, Département Médico-Universitaire d'Addictologie et Psychiatrie (DMU IMPACT), INSERM U955, IMRB, translational Neuropsychiatry lab, Fondation FondaMental, F-94010 Creteil, France. [20]Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust, Newcastle, UK. [21]Department of Psychiatry, University of Toronto, Toronto, ON, Canada. [22]Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON, Canada. [23]Mood Disorders Psychopharmacology Unit, University Health Network, Toronto, ON, Canada. [24]Brain and Cognition Discovery Foundation, Toronto, ON, Canada. [25]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, 68159 Mannheim, Germany. [26]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA 30322, USA. [27]Department of Psychiatry, University of Texas at Austin, Dell Medical School, Austin, TX, USA. [28]Department for Psychiatry and Psychotherapy, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Hauptstrasse 5, 79104 Freiburg, Germany. [29]Centre for Neuropsychopharmacology, Division of Psychiatry, Imperial College, London, London, UK. [30]Istituto di Neuroscience, University of Florence, Italy; Albert Einstein College of Medicine, New York, USA. [31]Department of Psychiatry and Behavioral Sciences, Psychiatry University of Miami, Miami, FL, USA. [32]Department of Biomedical, Metabolic & Neural Sciences, University of Modena, Modena, Italy. [33]VeraSci, Durham, NC, USA. [34]Department of Psychiatry, Amsterdam UMC, Vrije Universiteit and GGZinGeest, Amsterdam, the Netherlands. [35]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [36]Department of Psychiatry, University of Alabama at Birmingham, Birmingham, AL, USA. [37]Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada. [38]Federal Institute for Drugs and Medical Devices (Bundesinstitut für Arzneimittel und Medizinprodukte, BfArM), Bonn, Germany. [39]Office of Therapies for Neurological and Psychiatric disorders, Human Medicines Division, European Medicines Agency, Amsterdam, the Netherlands. [40]Biostatistics, GlaxoSmithKline, London, UK. [41]Laughren Psychopharm Consulting, LLC, Rockville, MD, USA. [42]GAMIAN-Europe (Global Alliance of Mental Illness Advocacy Networks-Europe), Brussels, Belgium. [43]Department of Psychiatry and Psychology, School for Mental Health and Neuroscience, EURON, Maastricht University Medical Centre, Maastricht, the Netherlands. [44]Faculty of Psychology, Center for Experimental and Learning Psychology, University of Leuven, Leuven, Belgium. [45]Data, Analytics and Methodology Taskforce, European Medicines Agency, Amsterdam, the Netherlands. [46]Department of Psychiatry, Hospital Universitari Vall d'Hebron (HUVH), Barcelona, Catalonia, Spain. [47]Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain. [48]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain. [49]Vita-Salute San Raffaele University, Milan, Italy. [50]Division of Neuroscience, Psychiatry and Clinical Psychobiology, IRCCS Scientific Institute Ospedale San Raffaele, Milan, Italy. [51]Strategic Projects Unit, Vall d 'Hebron Research Institute (VHIR), Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain. [52]Charité – Universitätsmedizin Berlin, Department of Psychiatry and Psychotherapy, Campus Benjamin Franklin, Berlin, Germany. [53]Janssen Research & Development, Beerse, Belgium. [54]Charité – Universitätsmedizin Berlin, Department of Psychosomatic Medicine, Campus Benjamin Franklin, Berlin, Germany. [55]University Medical Center Hamburg-Eppendorf, Institute of Neuroimmunology and Multiple Sclerosis (INIMS), Center for Molecular Neurobiology, Hamburg, Germany. [56]Department of Psychiatry, Erasmus University Medical Centre, Rotterdam, the Netherlands. [57]Medical Director Neuroscience for Region Europe at Novartis Pharma, Basel, Switzerland. [58]Department of Psychiatry, University of Tartu, Tartu, Estonia. [59]Faculty of Medicine, Department of Medicine, Centre for Neuropsychopharmacology, Division of Brain Sciences, Imperial College London, London, UK. [60]Documental Ltd, Tallin, Estonia; West Tallinn Central Hospital, Tallinn, Estonia. [61]Department of Global Regulatory Affairs, Neuroscience, Janssen Research & Development, LLC, Titusville, NJ, USA. [62]Group of Psychiatry, Mental Health and Addictions, Vall d'Hebron Research Institute (VHIR), Barcelona, Catalonia, Spain. [63]Universidad Autónoma de Barcelona, Barcelona, Spain. [64]Biomedical Network Research Centre on Mental Health (CIBERSAM), Barcelona, Catalonia, Spain. [65]Experimental Medicine, Janssen Research & Development, Janssen Pharmaceutica NV, Beerse, Belgium. [66]European Infrastructure for Translational Medicine (EATRIS), Amsterdam, Netherlands. [67]Neuroscience, Global Drug Development, Novartis Pharma AG, Basel, Switzerland. [68]Department of Neuroscience, Janssen Research & Development, LLC, Titusville, NJ, USA. ✉email: luca.sforzini@kcl.ac.uk