D4.1 Report on Consensus Disease Definition for TRD and PRD

"A Delphi-method-based consensus guideline for definition of Treatment-Resistant Depression for clinical trials"

853966 – EU-PEARL: EU Patient-cEntric clinical tRial pLatforms

WP4 – Integrated Research Platform for Major Depressive Disorder (MDD)

Lead contributors	Luca Sforzini; Courtney Worrell; Carmine M. Pariante (6 – King's College London, KCL)
	luca.sforzini@kcl.ac.uk; courtney.worrell@kcl.ac.uk; carmine.pariante@kcl.ac.uk
Other EU-PEARL contributors	Melisa Kose (6 – KCL); J. Antoni Ramos-Quiroga, Gara Arteaga-Henríquez (1 – VHIR); Edwin van de Ketterij (2 – EATRIS); Francesco Benedetti (7 – USR); Witte J. G. Hoogendijk (8 – EMC); Stefan M. Gold, Christian Otte (10 – Charité); Eduard Maron (22 – Documental); Adam J. Savitz, Heddie Martynowicz, Mark E. Schmidt, Katherine Woo (24 – Janssen); Yanina Flossbach (25 – Novartis)
	(Annex V: Full list of authors and affiliations; including external experts)
Reviewers	Woo Ri Chae (10 – Charité); Valeria Jordan Mondragon (25 –Novartis); Gabriela Perez-Fuentes (1 – VHIR)

Due date	30/04/2021
Delivery date	30/04/2021
Deliverable type	R
Dissemination level	PU

Description of Work	Version	Date
	V1.2	30/04/2021

Table of Contents

Document History

Version	Date	Description
V0.1	13/03/2020	First Draft
	22/05/2020	First online meeting
	01/06/2020	Comments (1st)
V0.2	07/07/2020	Second draft
	31/08/2020	Comments (2nd)
	09/10/2020	Second online meeting
	01/12/2020	Comments (3rd)
V1.1	10/12/2020	Third draft
	31/01/2021	Comments (4th)
	05/03/2021	Version sent to peer reviewers for feedback
	01/04/2021	Version updated with reviewers' comments, adjusted to Deliverable template and shared with PMO for review
V1.2	30/04/2021	Final Version

Definitions

- **Major depressive disorder (MDD).** Presence of MDEs (single or recurrent). The symptoms must not be a result of substance abuse or another medical or psychiatric conditions.

- **Major depressive episode (MDE).** Presence of depressive symptoms (according to diagnostic criteria) for at least two weeks. These symptoms must cause the individual clinically significant distress or impairment in social, occupational, or other important areas of functioning. The symptoms must also not be a result of substance abuse or another medical condition.

Abbreviations

Acronym / Abbreviation	Meaning
ATHF	Antidepressant Treatment History Form
ATRQ	Antidepressant Treatment Response Questionnaire
BAP	British Association for Psychopharmacology
D	Deliverable
DNA	Deoxyribonucleic acid
DBS	Deep brain stimulation
DSM	Diagnostic and Statistical Manual of Mental Disorders
DTD	Difficult-to-treat depression
CBASP	Cognitive behavioural analysis system of psychotherapy
CBT	Cognitive behavioural therapy
EMA	European Medicines Agency
EU-PEARL	EU Patient-cEntric clinicAl tRial pLatforms
FDA	Food and Drug Administration
HAMD	Hamilton Depression Rating Scale
ICD	International Classification of Diseases
IMI	the Innovative Medicines Initiative
IPT	Interpersonal therapy
IRP	Integrated research platform
MADRS	Montgomery-Asberg Depression Rating Scale
MDD	Major depressive disorder
MDE	Major depressive episode
MeSH	Medical Subject Headings
MINI	Mini-International Neuropsychiatric Interview
mRNA	Messenger ribonucleic acid
MTR	Multi-therapy-resistant
NbN	Neuroscience-based Nomenclature
NICE	National Institute for Health and Care Excellence
NIH	National Institutes of Health
PRD	Partially responsive depression
PROs	Patient-reported outcomes
QIDS	Quick Inventory of Depressive Symptomatology
SCID	Structured Clinical Interview for DSM
STAR*D	Sequenced Treatment Alternatives to Relieve Depression
TRD	Treatment-resistant depression
VNS	Vagus nerve stimulation
WHO	World Health Organization
WP	Work package

Abstract

Operational criteria for treatment-resistant depression (TRD) and partially responsive depression (PRD) as subtypes of major depressive disorder (MDD) are not unequivocally defined, with dissimilarities among different studies. In the present document we aim to use a Delphi-method-based consensus approach to define TRD and PRD and to deliver operational criteria for regulatory clinical trials, especially for Phase II/Proof of Concept. We systematically reviewed the literature and brought together a group of international experts (including clinicians, academicians, researchers, and regulatory bodies representatives) and other stakeholders. We discussed the current state-of-the-art and the main controversies regarding the current classification. Further, we provided recommendations on how to design clinical trials in the immediate future, and on how to guide research in unmet needs and knowledge gaps. This report will feed into the final and main objective 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 antidepressant treatments in TRD/PRD which could be used for regulatory protocols.

Keywords

Antidepressants, Depression, Diagnosis, Guidelines, Major depressive disorder (MDD), Partially responsive depression (PRD), Regulatory Science, Refractory depression, Treatment-resistant depression (TRD).

Summary

In the present report, we have systematically reviewed and summarized the evidence on treatment response in major depressive disorder (MDD) and elaborated a Delphi-method-based consensus document on treatment-resistant depression (TRD) and partially responsive depression (PRD). Our aim is to establish an explicit consensus on defining treatment resistance in MDD and to deliver operational criteria for randomized clinical trials, especially for Phase II/Proof of Concept clinical trials for new medications (or, potentially, repurposed medications) for TRD and PRD *(paragraphs 1.1 and 1.3)*.

We have identified and discussed the core gaps in knowledge and the main uncertainties of currently used TRD and PRD definitions, highlighting the absence of clear definitions and, importantly, a discrepancy between the most commonly used definitions and the most frequent inclusion criteria for clinical trials. This document has been developed as a part of the EUropean Patient-cEntric clinicAl tRial pLatforms (EU-PEARL) project, a public-private strategic partnership funded by the Innovative Medicines Initiative (IMI) to conceptualize and lead the design of an integrated research platform (IRP) *(paragraph 1.2)*.

The Delphi-based process involved the preparation and discussion of multiple draft documents, with continuous input from many experts and stakeholders. In the first draft of the present report, we reviewed the current literature on TRD/PRD definitions and inclusion/exclusion criteria for clinical trials in these indications. We then asked a number of international experts, all key leaders in TRD, to comment on the document and to participate to a virtual consensus meeting, which was held on the 22nd of May 2020. During the meeting, we debated the main controversies which had emerged from the literature review and the experts' comments. The next version of the document was then drafted including all the previous feedbacks, and additionally integrated with the discussion with stakeholders from the pharmaceutical and regulatory sectors, and patients' representatives, arising from a subsequent stakeholder meeting held on the 9th of October 2020 and additional stakeholders' written feedback. We then finalized the document performing a new literature review and merging all the contributors' comments *(paragraph 2)*.

We have structured the results section *(paragraph 3)* with different consensus recommendations on every point of debate, for each one specifying the level of consensus (strong, moderate or weak, indicating, respectively, near unanimity, substantial majority, or a small majority, of experts and stakeholders, supporting the statement). Here, we have summarized the main recommendations made throughout the report *(see also Table 2).*

1) TRD and PRD definitions

- A definition of TRD for clinical trials conducted for regulatory purposes is necessary. *Level of consensus – Strong*
- A definition of PRD – as a distinct group from TRD – for clinical trials conducted for regulatory purposes is recommended. *Level of consensus – Moderate*

2) Previous failed antidepressant treatments

- Failing at least two prior antidepressant treatments – in the current episode – is necessary to define TRD. *Level of consensus: Strong*
- A partial response to one single antidepressant treatment – in the current episode – is sufficient to define PRD. *Level of consensus: Moderate*
- There is no maximum number of previous antidepressant treatments for patients recommended to be included in TRD (and PRD) regulatory trials, but all documentable life-time treatment response should be recorded to define the TRD (and PRD) stage of a patient (see also below). *Level of consensus: Moderate*
- TRD (and PRD) definition (and differentiation) should be based on the current depressive episode only, and in the past two years only; if the current episode has lasted more than two years, treatments prior to the last two years should not be considered. *Level of consensus: Weak*
- To define TRD (and PRD), response to previous antidepressant treatments (within the current episode and in the past two years) can be ascertained retrospectively using structured interviews and clinical documentation. *Level of consensus: Moderate*

3) Type, dose, and duration of an 'adequate' antidepressant treatment trial

- To define TRD, the two different treatment failures must involve two established medications for MDD, with different mechanisms of action. *Level of consensus: Moderate*
- Regulatory clinical trials for TRD (and PRD) may include patients who failed to respond (or partially responded) to augmentation/combination treatment strategies, but these need to be primarily based on medical records. *Level of consensus: Strong*
- Regulatory clinical trials for TRD (and PRD) may include patients who failed to respond (or partially responded) to brain stimulation treatments, such as transcranial magnetic stimulation (TMS) and electroconvulsive therapy (ECT). *Level of consensus: Moderate*
- Regulatory clinical trials for TRD (and PRD) should not include patients who failed to respond (or partially responded) to vagus nerve stimulation (VNS) and deep brain stimulation (DBS). *Level of consensus: Moderate*
- Regulatory clinical trials for TRD (and PRD) may include patients who failed to respond (or partially responded) to structured psychotherapy. *Level of consensus: Strong*
- The minimum effective dose of a medication indicated in MDD is enough to *define a treatment failure for the purpose of establishing* TRD (and PRD). *Level of consensus: Moderate*
- For conventional medications indicated in MDD, a treatment given for at least four weeks in duration is sufficient to define *a treatment failure for the purpose of establishing* TRD (and PRD). *Level of consensus: Moderate*
- Patients' discontinuation of treatment before the completion of the fourth week, should not be considered as *a treatment failure for the purpose of establishing* TRD (and PRD). *Level of consensus: Strong*

4) Clinical presentation of TRD and PRD patients

- No specific types of symptoms of MDD should either be prerequisite, or excluded, from the definition of TRD (and PRD), but symptoms and specifiers should always be recorded. *Level of consensus: Strong*

- Patients with bipolar depression should be excluded from TRD (and PRD) studies, as this is a separate condition from unipolar depression (MDD). *Level of consensus: Strong*
- Patients with comorbid personality disorders or other mental disorders should be excluded from TRD (and PRD) studies only when their onset is properly documented as independent and antecedent to the MDD diagnosis. *Level of consensus: Moderate*
- Patients with comorbid substance use disorder that is active and severe should always be excluded from TRD (and PRD) studies, independently from the onset; *in contrast*, patients with comorbid substance use disorder that is active and mild/moderate should be excluded from TRD (and PRD) studies only when the onset is properly documented as independent and antecedent to the MDD diagnosis. *Level of consensus: Moderate*

5) Diagnostic tools and measures of outcome

- Maudsley Staging Model is the suggested instrument to define the degree of treatment resistance historically. *Level of consensus: Moderate*
- Clinician administered MADRS10 is the suggested outcome instrument to assess treatment response (and remission) and, together with patient-reported QIDS-SR, can be used to assess TRD and PRD status. *Level of consensus: Moderate*
- Criteria for remission, response, and partial response should not be relaxed in regulatory clinical trials for TRD (and PRD); shorter versions of the traditional scales, such as the HAMD6 and the MADRS6, should not be currently preferred to traditional scales, although may become more relevant in the future with fast-acting interventions. *Level of consensus: Moderate*

6) Future directions

- Future research should be more patient-centred, recognizing, and targeting different clinical phenotypes of TRD and PRD underpinned by a specific biological mechanism. *Level of consensus: Strong*
- 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 patient. *Level of consensus: Strong*
- Currently, no biomarker has been validated in clinical practice or in clinical trials to identify TRD (and PRD) patients, or to further stratify them; however, collection of biological samples for subsequent subgroup or stratified analyses is recommended. *Level of consensus: Moderate*
- Patients' preferences, perspectives, and reported outcomes should be included in future TRD (and PRD) diagnostic tools and outcome measures. *Level of consensus: Strong*
- 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. *Level of consensus: Moderate*

Standardised research for regulatory purposes is the only way to advance the MDD field towards tailored treatments. This consensus document aims to fill the gaps in knowledge within TRD and PRD research, providing straightforward and replicable criteria that we hope may help to set the stage for present and future clinical (and platform) trials.

1. Introduction

1.1. Core gaps in knowledge and uncertainties of currently used MDD, TRD and PRD definitions

1.1.1. Major Depressive Disorder (MDD)

Major depressive disorder (MDD) is a complex and heterogeneous clinical condition, as outlined in the most commonly used diagnostic criteria, the fifth edition of the American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders (DSM-5) ([1]), and the eleventh version of the World Health Organization International Classification of Diseases (ICD-11) ([2]). They both describe clinical features to diagnose MDD, including alterations in mood, interests, and pleasure, as well as impairments in cognition, and other biological functions such as sleep, appetite, and sex drive ([3]). According to the World Health Organization (WHO), more than 264 million people were affected by MDD in 2017, with a rising trajectory estimated at >18% increase spanning the ten-year period between years 2005-2015 ([4]).

Major depressive episodes (MDE) can be either single or recurrent. Based on the duration of symptoms, two weeks (in both the DSM-5 and the ICD-11) are considered the minimum required duration of symptoms for a defined diagnosis, but the classifications also include forms of 'brief depression', if the episode lasts less than two weeks, or 'persistent depression' (or 'dysthymia'), when the episode lasts for more than two years. These can all be qualitatively similar, and distinct only based on temporal criteria. Although the clinical course of MDD may vary widely, the majority of patients (i.e., >75%) develop recurrent episodes, with recurrences usually within two years of recovery, that is, after a period of 'sustained remission' *(as we will describe later in 3.2 and 3.5)* ([5]; [6]). All the evidence abundantly demonstrates how depression impacts the lives of those who are affected ([7]). MDD is one of the leading contributors to the global burden of disease, being one of the three leading causes of all-age years lived with disability worldwide ([4]). In addition to the human capital costs and decrements in quality of life, MDD is also associated with increased medical burden, suicidal behaviour, and all-cause morbidity and mortality ([8]; [9]; [10]; [11]).

There are several specifiers of depression based on clinical features, regardless of the duration, as described in the DSM-5. For example, depression 'with melancholic features' is characterized by the main symptoms of hopelessness and lack of reactivity of mood, combined with symptoms regularly worse in the morning, early-morning awakening, marked psychomotor alterations, significant anorexia or weight loss, excessive guilt, and a distinct quality of the depressed mood described as profound despondency, despair, moroseness, or 'empty-mood'. MDD can also present 'with atypical features', characterised by reactivity of mood (e.g., mood brightens in response to actual or potential positive events), often together with reverse neurovegetative symptoms (hypersomnia and hyperphagia), a long-standing pattern of interpersonal rejection sensitivity, and so-called 'leaden limb paralysis', or heaviness in upper and lower limbs together with profound fatigue. One of the MDD specifiers most frequently observed is 'with anxious distress', when symptoms of anxiety are present during the majority of days of the MDE ([12]); this is relevant as the presence of anxiety can make the depression more difficult to treat ([13]). MDD may also be characterized as having 'mixed features', when manic/hypomanic symptoms (not meeting criteria for mania or hypomania) co-occur during the MDE;

this specifier has also been associated with antidepressant resistance, and it is possible in MDD, even if it is more frequently associated to a bipolar disorder, thus a differential diagnosis is crucial ([14]; [15]). Finally, depression is characterised as being with 'psychotic features' in the presence of psychotic symptoms, which can be, more frequently, mood-congruent (e.g., typical depressive themes of personal inadequacy, guilt, disease, death, nihilism, or deserved punishment), or mood-incongruent. The pathogenesis, phenomenology, phenotype, and illness trajectory in MDD are highly heterogenous, inviting the need for a more personalized approached to treatment selection in individual patients ([16]). However, there is still a lack of individualised or personalised treatments readily available to those in need ([17]). In fact, despite initial attempts to shift this paradigm, and even if there are many international clinical guidelines directing treatment protocols (also based on the presence of specific symptoms), MDD treatment at an individual level is often still entrapped in a one-size-fits-all, trial and error approach.

1.1.2. Inadequate response to treatment, treatment-resistant depression (TRD) and partially responsive depression (PRD): the quest for definitions

More than two dozen medications have proven efficacy in the treatment of MDD ([18]), but inadequate response to treatment, even with multiple medication exposures, remains a common clinical scenario, representing an ongoing clinical challenge ([19]; [20]; [21]; [22]; [23]). It is estimated that about one third of MDD patients (ranging from 20% up to more than 50%) do not achieve full symptomatic remission after one monoamine antidepressant treatment ([24]; [25]; [19]), and up to 15% remain unwell after multiple treatments ([26]). This percentage of patients who do not respond further increases when we also consider the functional remission together with the symptomatic one; in one study, combined symptomatic and functional remission was achieved by 23% of MDD patients after one antidepressant treatment, while symptomatic remission only was achieved by 38% ([27]). Moreover, patients with inadequate responses to initial treatments, even if they finally respond to a successive treatment, will have a very high overall relapse rate while continuing the treatment (65% after 2 failed treatments, and 71% after 3) ([19]). As mentioned above, there are several clinical factors associated with lack of response to medications for MDD ([28]), such as comorbid anxiety and current suicidal risk ([29]), or early onset of MDD ([30]).

Individuals with inadequate responses can show either no response, or partial response, to medications for MDD. This has led to the introduction of different terms, such as treatment-resistant depression (or TRD) or partially responsive depression (or PRD) – two terms that are discussed and defined at length in this report. Moreover, 'inadequate response to antidepressant treatment' is not an all-or-nothing phenomenon, but rather a continuum that ranges from PRD, to TRD, to 'multi-therapy-resistant MDD (MTR-MDD)' ([31]), to 'refractory depression' which implies an absence of response to all currently available treatments administered over a prolonged period of time. Unfortunately, there is a lack of consensus definitions around concepts such as response, PRD, TRD, and adequate dose and duration of antidepressant treatment ([32]). Even when considering the treatment of TRD/PRD, such as pharmacological augmentation, recommendations are not consistent across current guidelines ([33]) and evidence of effectiveness is sparse ([34]). This ambiguity complicates the generalizability of results to the real-world setting, and profoundly hinders research and progress in this field. A serious consequence is that there is no uniform population for clinical studies on PRD or TRD, for clinical and biological investigations, and especially for regulatory Phase II/Proof of Concept or Phase III clinical trials for new medications (or, potentially, repurposed medications) for TRD and PRD. Thus, a better and possibly clearer definition of clinical populations to be included in regulatory

trials in TRD and PRD is at the core of the present document. These definitions should not be taken to indicate a subtype or type of depression but pragmatic criteria to enable standardisation of clinical trials. Importantly, and as extensively discussed below, regulators acknowledge that response, partial response, and nonresponse exist on a continuum without universally accepted definitions, but nevertheless distinguish between these conditions (MDD, PRD, TRD) when they consider the types of studies they require to demonstrate efficacy in depression. Indeed, treatments for TRD and adjunctive treatment of MDD (for PRD) are already accepted regulatory paths for drug approval, even if with no definitive evidence and still ongoing debate among academicians, clinicians, and regulators *(see 3.1, Issues of Debate)*.

It is also important to note that, in a recent consensus statement, a group of experts (some of them participating to the current report) have suggested that both the terms PRD and TRD are semantically and operationally not ideal, and thus have proposed to use the broader concepts of 'difficult-to-treat depression (DTD)' or 'suspected DTD' ([23]). The authors caution that this definition may not be specific and objective enough to define clinical populations for regulatory trial objectives – something which, as we have said, is at the core of the present report. They described DTD as "depression that continues to cause significant burden despite usual treatment efforts". This concept obviously overlaps highly with PRD and TRD, but still with some important differences. The notion of DTD is clinically interesting, and introduces a more flexible, multidimensional and longitudinal definition, and as such is discussed amply in this report. Notably, the authors themselves acknowledge that, "what constitutes significant burden in the definition of DTD is subjective and likely to vary between patients", thus implicitly involving the patient's point of view, but also derogating from an operational definition of DTD. This is also likely to vary among clinicians and raters, as to what degree of reported patient burden they see as 'significant'. However, notwithstanding these limitations, the concept of DTD is important in order to highlight the complexity, the limitations, and the multiple open questions that are intrinsic in any attempts to operationalise 'inadequate treatment-response' to medications – a topic that we also discuss extensively in this report.

1.2. The EU-PEARL project

This initiative is part of a broader European project, the EU Patient-cEntric clinicAl tRial pLatforms (EU-PEARL). This is 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, through a network of investigational sites, with a federated and shared data platform and agreed regulatory pathway. The overall program will focus on four different diseases: MDD, tuberculosis, non-alcoholic steatohepatitis, and neurofibromatosis. Thirty-six institutions, including European university hospitals, research centres, patient groups, non-profit product developers and pharmaceutical companies, are cooperating in this consortium. The final objective is to shape future medication development through a systematic approach that enables cross-company collaborative platform trials that are patient-driven by design and patient-centred by outcome. One work package specifically concentrates on MDD (i.e. WP4), particularly for patients who have failed to respond adequately to first line antidepressant treatment(s), and this report has been produced as part of this work package. More information on the project can be found on the official website ([35]).

The use of IRPs could help overcome several of the current limitations in the field, both from a clinical as well as an economic perspective. To date, clinical trials typically evaluated the efficacy of one treatment at a time. This methodology implies long and sequential cycles to develop a specific medication, with substantial economic investments and delays in getting the most effective treatments to patients. Moreover, competing trials frequently pose a challenge for patient recruitment, and patients often struggle to navigate the complex trial landscape to find the clinical study best suited to them. A potential strategy to overcome these issues is the use of platform trials. These are clinical trials with a single master protocol in which multiple treatments are evaluated simultaneously and/or sequentially ([36]; [37]). Importantly, they can have an adaptive design – with flexible features – allowing the removal of treatments or groups (for example, for futility) or the addition of new groups or treatments during the course of the trial ([36]; [38]). Since there are many medications for the first line treatment of MDD ([18]), this project focuses on patients who do not respond adequately to initial treatment(s) (TRD and PRD), with the aim of developing methodology to demonstrate preliminary effectiveness of new medications (proof of concept) for this populations. EU-PEARL project will, for the first-time, allow adaptive platform trials to focus on TRD/PRD. In addition to developing an IRP for MDD, the MDD work package of EU-PEARL also aims to develop the protocol for a longitudinal natural history study, to better understand the course of MDD as well as to develop a prospectively-identified and recruitment-ready cohort for IRP studies.

1.3. Aim of the report

As stated previously, the ambition of the WP4 EU-PEARL programme of research is to establish a Phase II/Proof of Concept protocol for platform trials of potential new medications for TRD and PRD for use in randomized clinical trials for new treatments and informing confirmatory trials. While this may seem too narrow of a focus for this report, the ambition is to make this document an important consensus statement on TRD and PRD, which both offers well delineated definitions for the present and clarifies the needs and possible procedures for the future, including research needs.

The aims of this report are:

1. Discuss and optimise TRD and PRD definitions that could be used at an individual patient-level with information that is currently routinely collected in clinical practice, including clinical interviews of current and retrospective symptoms and treatments, and hospital-, pharmacy- and other electronic health data;
2. Make firm recommendations for measures and variables that should be included in future clinical trials for TRD/PRD, including measures of outcomes, and in longitudinal cohorts or electronic health records, so that the proposed definition of TRD/PRD could be further refined in the future with input from regulatory experts; and, finally,
3. Indicate which are the important areas of uncertainty that require further research, including broader theoretical issues and specific research questions.

While providing consensus statements on these three issues, we will also express the level of agreement for each specific recommendation and discuss their limitations. Finally, throughout the document we will balance the need of regulatory authorities with those of clinicians and scientists, and describe the difference when indicated. This document will fit as a deliverable for the WP4 EU-PEARL programme. A shorter version of it will be submitted for open-access publication.

2. Methodology underpinning this report

2.1. Delphi-method-based consensus document and consensus meetings

We have decided to structure our consensus document following a Delphi approach. This method is particularly useful to gain consensus where an investigational approach is not possible. In summary, it relies on a panel of experts through several rounds of anonymized feedback on progressively updated versions of a report/questionnaire ([39]). This method has been recommended as the method of choice for developing reporting guidelines in health research in a recent guidance ([40]). Notably, in this same guidance, the authors endorsed the use of face-to-face consensus meetings together with a Delphi approach, as we did in the present manuscript *(see below)* ([40]). Despite its wide application in medical and psychiatric research ([41]), a recent systematic review found only 25% of the reporting guidelines used the Delphi method to develop consensus ([42]), advocating the use of this methodology.

In order to gather experts' opinion on these gaps in knowledge, we invited experts with a track record of publications in this area or stakeholders with clear expertise, from a broad range of specialties and diverse experiences in clinical practice, academia, industry, and regulatory agencies, including clinicians and scientists associated with the EU-PEARL project (see Annex V: Full list of contributors to D4.1 and acknowledgement).

As a first step, we developed a draft document on the up-to-date literature on current TRD/PRD definitions and inclusion/exclusion criteria for regulatory clinical trials and for clinical trials for new treatments in these indications, with a questionnaire at the end to gather opinions on the most debated issues *(questionnaire and list of contributors in*
and Annex V: Full list of contributors to D4.1 and acknowledgement, respectively).

After we received experts' comments and answers, we then organized a first consensus meeting by video conference, conducted on the 22nd of May 2020 *(consensus meeting agenda in the Annex II),* with international experts and EU-PEARL members. Here, the experts gave presentations focused on key points of uncertainty. Each session was followed by a discussion within the whole group about the requirements for future regulatory studies.

We merged all the comments and answers to the first draft with those arisen from the meeting; the resulting draft was circulated again to all the contributors for their feedback. The resulting document (with an updated literature review) was circulated to a group of stakeholders, including regulatory authorities and patients' representatives, and discussed in a second online consensus meeting on the 9th of October 2020, with stakeholders and EU-PEARL members, with additional stakeholders providing written feedback *(see Annex II)*.

The final version was then circulated and approved by all authors. Contributors' identity was revealed, also to allow face-to-face meetings, however, the different versions of the document and summary of previous feedback were always anonymized.

2.2. Evidence base and review of the literature

In order to develop the initial draft and start the discussion, we performed a comprehensive review of recent systematic reviews, consensus papers, and regulatory documents on TRD/PRD, up to the end of February 2020. A subsequent search was performed after the completion of the report in January 2021.

Using MEDLINE PubMed® we searched for the following MeSH terms and keywords: ("treatment-resistant depression" or TRD), ("partially responsive depression" or PRD), and ("major depressive disorder" or MDD or "unipolar depression" and "response to treatment"). We have further filtered results including only systematic reviews, meta-analyses, books, and documents. In addition, we searched for consensus statements, management guidelines, and regulatory documents through other website and platforms, including British Association for Psychopharmacology (BAP), European Medicines Agency (EMA), Food and Drug Administration (FDA), National Institute for Health and Care Excellence (NICE), National Institutes of Health (NIH), and World Federation of Societies of Biological Psychiatry (WFSBP). We then summarized evidence from the most recent papers. More specifically, these included the reviews by Gaynes et al. ([21]), Salloum and Papakostas ([22]) and McAllister-Williams et al. ([31]), as well as the consensus statement by McAllister-Williams et al. ([23]). In particular, the review by Gaynes et al. ([21]) comprises 185 unique studies in TRD; the one by Salloum and Papakostas ([22]) discusses 18 articles with proposed TRD definition and staging model, and empirical work to support the model/definition; McAllister-Williams et al. ([31]) discuss a concept of multi-therapy resistant depression, defining patients well in excess of a TRD definition and representing a potential 'upper limit' for inclusion in this kind of studies; and the consensus statement ([23]) discusses the concept of DTD, as mentioned above. Key documents from regulatory authorities included FDA draft guidance ([43]) and the EMA guideline ([44]) on depression *(Table 1).*

To complete the final version of the document, before final approval of all the contributors, we performed a systematic review to identify any relevant paper published in the last few months. Using MEDLINE PubMed® database, we searched for reviews, systematic reviews, meta-analyses, and guidelines published from the 1st of March 2020 to the 22nd of January 2021, using the following search string: ((Treatment-resistant depression) or (TRD) or (partially responsive depression) or (PRD) or (difficult-to-treat depression) or (DTD)) and ((definition) or (diagnosis) or (criteria)). The electronic search returned 50 records, 20 of which specifically evaluating MDD non-responder patients. Only one article – besides the aforementioned statement by McAllister-Williams et al. ([23]) – provided a rationale for TRD or PRD definitions ([45]). In this review, the authors stressed the importance of distinguishing between inadequate response (PRD) and non-response (TRD), presenting the same definitions and the same cut-offs we endorse in our document, further corroborating our recommendations *(see point 3.1)* ([45]). In order to develop the initial draft and start the discussion, we performed a comprehensive review of recent systematic reviews, consensus papers, and regulatory documents on TRD/PRD, up to the end of February 2020.

2.3. Structure of the report

In the core part of the present report, the *'Results and consensus recommendations'*, we report the strongest view, that is, the view supported by the largest number of the experts. At the end of each

section, we provide consensus recommendations for that specific topic or issue (listed as bullet points). However, we also highlight the most important areas of uncertainty inferred from the literature or from the further discussion in the online meetings, and we also discuss the 'level of consensus' (strong, moderate, or weak) on each specific point. In the first five sections *(from 3.1 to 3.5)*, we have provided key recommendations which can be immediately implemented, within current research practice, discussing the main 'issues of debate' for each section, while in *section 3.6* we have discussed specific needs and directions for future research. We have summarised all the recommendations with their respective level of consensus in *Table 2*.

2.4. General limitations of the report

We have to acknowledge that the recommendations in this report are based on the experts' (and stakeholders') views and consent, and not on any hard piece of research or clinical evidence. As mentioned above, specific open questions and points for debate or further research are mentioned in each section, but we wanted to highlight here the general theoretical limitations of this report. While doing this, we also want to emphasise that all experts agreed for the need of these consensus-based definitions proposed here, in order to improve the future evidence-base. For specific recommendations we present here only statements where a majority agreement was expressed among the experts and stakeholders, although this agreement was at times strong (unanimous, or almost unanimous), at times moderate (a substantial majority) and at times weak (a small majority).

Perhaps the most important conceptual limitation is that we cannot exclude that significant phenotypic (and potentially biological) heterogeneity is present in the defined samples, even if our aim is to create definitions that identify 'clinically homogenous' samples, based on a pragmatic and non-aetiologically based approach. Future research in these homogeneous samples will allow to define exactly this: the inherent variability, if any, in clinical features, biomarkers, and clinical response. Second, the opposite risk is also present: that we define groups so narrowly that the findings are not generalisable to the larger population of depressed patients. This is important, as the aim of the initiative is not just about identifying specific groups of MDD patients for research studies, but to produce data that are representative of patients in clinical settings, finally translating into better care. In order to avoid definitions that are too intrinsically narrow, in the present report we continue to identify patients based on diagnostic criteria, as reported in the DSM-5 and ICD-11, thus including all MDD patients regardless of their clinical presentation; exceptions are MDD patients with particular comorbidities, who fail certain brain stimulation techniques, and who are aged below 18 and over 65 years, who will require to be studied separately *(see sections 3.3.1 and 3.4)*. Moreover, to fully understand 'who are' the patients that are defined by a specific TRD (or PRD) definition, and where do they stand on the tension between heterogeneity (within the samples, and potentially across samples) and homogeneity (and potential lack of generalisability), we recommend the type of information that we need to collect in future studies employing these definitions.

It is also important to emphasise that TRD and PRD concepts are often based upon a conceptualisation of depression as being an episodic illness with good inter-episode recovery; this is however not the case for several patients, who often have a chronic illness with waxing and waning course. Merging these two types of patients into a single TRD (and PRD) definition has some intrinsic problems, for example when trying to identify the beginning of the current episode in order to define

response to current and recent antidepressants. However, we recognise the disease continuum between TRD, PRD, and response to treatment, and that, over time, and depending on response to treatments, a subject may fall into different disease states. We therefore want to emphasise that *at any one time* a patient who is not a responder necessarily falls into one of these two disease states, TRD or PRD. Our aim is to help defining these two putative disease states as much as possible, while acknowledging all the limitations of this approach. More limitations for different topics will be discussed separately in the *Issues of debate* sections.

3. Results and consensus recommendations

3.1. TRD and PRD definitions

3.1.1. What is lack of response?

A treatment response (or 'complete response') in MDD is a response to a treatment defined by a reduction of at least 50% in MDD severity *(this definition will be further discussed in section 3.5)*. What is a 'lack of response', encompassing both TRD and PRD, is thus the mirror image of this definition, that is, a less than 50% reduction in depression severity, with TRD further associated, in most definitions, with lack of response to at least *two* medications at an adequate dose and duration.

There is, however, much variability around the definition of TRD, as is evidenced through TRD clinical trials. In a recent review, Gaynes and colleagues ([21]) firstly performed a narrative literature review on the most commonly used TRD definitions, comprised of systematic reviews and guidelines or consensus statements. Following this, they systematically reviewed whether those definitions coincided with the inclusion criteria for patients in TRD studies. Surprisingly, they found that only 37% of intervention studies in TRD had enrolled patients meeting the most recommended criteria (at least two failed antidepressant treatments of adequate dose and duration). In fact, the most common definition for inclusion in TRD intervention trials involved a minimum of only one previous failed treatment (48%), though it was not always clear if this failure was in a past or current episode *(we will discuss these specific issues in sections 3.2 and 3.3)*. When the authors additionally considered adequate dose and duration of treatments *(points 3.3.2, 3.3.3)*, they reported that only 19% of studies enrolled patients fulfilling all the criteria. This is particularly relevant because it suggests that the majority of studies on TRD do not use the most recommended definition described above, and therefore the data across these TRD studies cannot be pooled or compared, as they evaluate potentially different populations. Moreover, studies in which patients were not treated with an adequate dose of an approved medication for MDD, and for an adequate duration, may consequentially include patients who do not truly meet the criteria for TRD *(see below)*.

3.1.2. Operational criteria for TRD and PRD

The use of the term 'PRD' for clinical and regulatory studies further complicates the situation, as both TRD and PRD tend to be used for patients that show 'lack of response'. Based on most papers in the literature and feedback from the experts and stakeholders, we propose here that TRD should be used to indicate a more narrow definition of patients who show *a reduction of less than 25% in depression severity using prospective psychometric assessments*, or an equally insufficient improvement in depression severity as assessed retrospectively through a clinical interview.

On the other hand, PRD should be used to indicate patients who show *a reduction of between 25% and <50% in depression severity using prospective psychometric assessments*, or by an equal improvement in depression severity as assessed retrospectively through a clinical interview. This defines TRD and PRD as two subgroups of non-complete responders (<50% improvement), independent from each other (<25% and 25–<50%, respectively), as supported by a number of guidelines and other expert documents ([46]; [24]; [47]; [45]).

Another important difference between TRD and PRD is that, while two different failed antidepressant treatments are necessary for the TRD definition, only one is sufficient to define PRD *(as we will further elaborate in section 3.2)*. The majority of the experts, even though with moderate consensus, highlighted the importance of this distinction between TRD and PRD for randomized clinical trials for new treatments, in terms of differentiating between true TRD patients (<25% improvement to *two* medications), and those who respond but not completely (PRD, 25-<50% improvement to *one* medication), mostly because of the potential advantage of subgrouping patients for different randomized controlled trials (for example, involving augmentation vs. switching*, see below*) rather than considering all non-responder MDD patients as the same *(see also next section and 3.4)*.

Issues of debate

While most experts endorsed the importance of developing a clear definition of TRD and PRD for clinical trials, some have argued that these definitions could be unhelpful from a clinical and a conceptual perspective, as they would be arbitrary to apply (defining an arbitrary threshold on a continuum) and influenced by different healthcare systems. The potential benefits and disadvantages of such a distinction were among the most highly debated topics of the present document.

The EMA guideline does highlight the discrepancy between everyday practice and the conceptual elaboration and definition of clear criteria for TRD and PRD, and indeed they highly recommend the generation of validated criteria ([44]) to address this. It is however unclear whether two distinct definitions are needed, or a single category (also on a continuum) would be more appropriate. In fact, some experts, also representatives of regulatory agencies, even when overtly supporting the need for TRD/PRD definitions, acknowledged the continuum between response, PRD, TRD, multi-therapy-resistant MDD, and refractory depression (as discussed in the *Introduction*), and that TRD/PRD should thus be better characterised using a dimensional, 'staging' approach (as discussed below in *section 3.6*). Of course, the regulatory implications of differentiating between TRD and PRD are quite clear, and the positive impact comes from the opportunity to test key clinical questions leading to different licensed treatments for different indications or clinical situations.

However, even for regulatory purposes these definitions create a potential confusion between a generic, broader definition of TRD (a reduction in depression severity between 0 and <50%, which includes PRD, viewing TRD and PRD as on a continuum) and a more stringent definition of TRD (<25% improvement)) to be distinguished from PRD (25–<50% improvement). Specifically, the stricter TRD definition could be used to enrol patients in randomized controlled trials for new medications, while PRD patients may be preferentially recruited for trials to study augmentation/add-on treatments ([44]; [47]; [43]).

However, this 'design-based' distinction is not entirely supported from literature evidence. As an example, in a pooled analysis of two studies examining the efficacy of adjunctive aripiprazole in MDD, patients with TRD seemed to have a larger treatment effect compared with those with PRD ([48]); or also, treatment with intranasal esketamine, which has been proven effective in augmentation to oral antidepressant therapy in TRD patients ([49]). While most experts endorsed the importance of developing a clear definition of TRD and PRD for clinical trials, some have argued that these

definitions could be unhelpful from a clinical and a conceptual perspective, as they would be arbitrary to apply (defining an arbitrary threshold on a continuum) and influenced by different healthcare systems. The potential benefits and disadvantages of such a distinction were among the most highly debated topic of the present document. The EMA guideline does highlight the discrepancy between everyday practice and the conceptual elaboration and definition of clear criteria for TRD and PRD, and indeed they highly recommend the generation of validated criteria ([44]) to address this. It is however unclear whether two distinct definitions are needed, or a single category (also on a continuum) would be more appropriate. In fact, some experts, also representatives of regulatory agencies, even when overtly supporting the need for TRD/PRD definitions, acknowledged the continuum between response, PRD, TRD, multi-therapy-resistant MDD, and refractory depression (as discussed in the *Introduction*), and that TRD/PRD should thus be better characterised using a dimensional, 'staging' approach (as discussed below in *section 3.6*). Of course, the regulatory implications of differentiating between TRD and PRD are quite clear, and the positive impact comes from the opportunity to test key clinical questions leading to different licensed treatments for different indications or clinical situations. However, even for regulatory purposes these definitions create a potential confusion between a generic, broader definition of TRD (a reduction in depression severity between 0 and <50%, which includes PRD, viewing TRD and PRD as on a continuum) and a more stringent definition of TRD (<25% improvement)) to be distinguished from PRD (25–<50% improvement). Specifically, the stricter TRD definition could be used to enrol patients in randomized controlled trials for new medications, while PRD patients may be preferentially recruited for trials to study augmentation/add-on treatments ([44]; [47]; [43]). However, this 'design-based' distinction is not entirely supported from literature evidence. As an example, in a pooled analysis of two studies examining the efficacy of adjunctive aripiprazole in MDD, patients with TRD seemed to have a larger treatment effect compared with those with PRD ([48]); or also, treatment with intranasal esketamine, which has been proven effective in augmentation to oral antidepressant therapy in TRD patients ([49]).

Taken all this issues into consideration, for the purpose of this report, we will define TRD in its stricter meaning, thus different from PRD, in order to define separate categories for clinical trials that intend to focus on one or the other of these types of patients (or also on both, with subsequent stratification). We therefore think that, despite all the uncertainties, this distinction may be useful to provide a more tailored classification of non-responder MDD patients, to pursue a 'precision medicine' with more tailored interventions.

Recommendations

- A definition of TRD for clinical trials conducted for regulatory purposes is necessary.
    - *Level of consensus – Strong*

- A definition of PRD – as a distinct group from TRD – for clinical trials conducted for regulatory purposes is recommended.
    - *Level of consensus – Moderate*

3.2. Previous antidepressant treatments

TRD and PRD definitions are built around the failure to respond to antidepressant treatments. Before discussing specific issues related to the treatment itself *(next section, 3.3)*, we will here focus broadly on how to investigate previous treatments. We will provide recommendations for (1) how many previous treatments should be considered, (2) in which episode (current and/or past) and (3) the way the treatment failure or partial response should be assessed for inclusion in clinical trials (prospectively and/or retrospectively).

3.2.1. Number of previous treatments

No agreement exists on how many previous treatment failures are required in order to define TRD (and PRD) in an identified patient. The vast majority of systematic reviews and guidelines or consensus statements concur in defining TRD as the presence of at least one or, more often, two, failed sequential treatment attempts at an adequate dose and for an adequate duration during the index episode. The FDA guidance ([43]) confirms that no universally accepted definitions or cut-offs exist for TRD or PRD; they propose that TRD studies should include patients who have not responded to *more than one* prior medication administered at an adequate dose and duration. Similarly, the EMA opts for the same definition when considering the matter "in a clinical pragmatic view" – at least *two failed treatments (Table 1)* – still highlighting that no validated clinical criteria and thresholds to define TRD and PRD are available, and supporting the generation of such data. This same definition of at least two failed medications is confirmed as the most commonly used in the review by Gaynes and colleagues ([21]). In their consensus statement, McAllister-Williams and colleagues ([23]) endorsed this position, suggesting that suspected DTD should normally be considered after patients have failed at least two adequate treatments. Nevertheless, under certain conditions – for example, where standard treatments are contraindicated – some patients might be considered to also have DTD with a single treatment failure *(Table 1)*. There is also an uncertainty around PRD, even if a partial response to a single antidepressant treatment is considered enough. EMA documents do not specify a precise number of previous treatments to diagnose PRD ([44]), while the FDA broadly indicates that "for adjunctive treatment, studies should include patients with partial responses to other antidepressant therapies" ([43]).

Consistent with the most commonly used criteria, the majority of our experts agreed on the fact that, for regulatory purposes, TRD should be defined after a minimum of *two* failed treatments with <25% of improvement with adequate dosing and duration, while PRD can be defined even after a single treatment (improvement 25-<50%) with adequate dosing and duration. We will further discuss in detail what an adequate antidepressant treatment is (type, dose, and duration) *(section 3.3)*.

We also discussed whether to set a limit on the maximum number of previous failed treatments. Some experts were in favour of defining a maximum number of previous treatments (mostly from three to five), as a maximum number of failed medications for MDD is an exclusion criterion in many clinical trials. However, the agreed consensus was *not* to exclude multiple-resistant patients, but rather to include this information – recorded both for the current episode and lifetime – as part of the staging (dimensional) approach, to capture the complexity and variability of the TRD population *(see also section 3.6)*.

3.2.2. Current or past episodes

Another conceptual and pragmatic issue is whether the evaluation of non-response or partial response should include prior episodes or be confined to the current/index episode. In general, the agreed definition of TRD for clinical trial purposes would normally include a current failure and a past failure at minimum (i.e., someone is currently receiving a medication indicated for MDD and they are still depressed according to current clinician's assessment; also they were treated with another one in the past, and it is assessed retrospectively that they did not respond); however, it is unclear whether both of these treatments should apply to the same (current) episode or to clearly distinct episodes. Although the EMA definition of MDD ([44]) 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.

The experts' majority opinion was that, in order to have a clear and reproducible definition for research and regulatory purposes, the definition of TRD should include two treatment failures (<25% improvement) within the *current* episode, and for PRD should include partial response (25-<50% improvement) to at least one treatment within the *current* episode. Of course, the focus on current episode does not reduce the importance of a correct assessment of the patient's treatment history that could further guide therapeutic decisions and choices. In addition, we all acknowledged the great difficulty in retrospectively defining the response to a past treatment, especially if the current episode is of long duration and the treatment was closer to the onset of the episode. To address these concerns, the additional recommendation (besides focusing on the current episode only) is that only treatment failures within the last two years should be considered, although it is important to stress that the consensus on this point was weak, with both shorter and longer periods proposed *(see Issues of Debate below)*.

3.2.3. Prospective or retrospective assessment of treatment failure or partial response

A separate, related issue, which is particularly important for regulatory clinical trials, is whether at least one treatment failure should be 'prospective within the trial' (i.e., the trial starts with an established medication for MDD, and then the person is offered a new intervention only if they fail to improve on an adequate dose of the conventional medication within an adequate period of time during the trial). This is an important consideration for a clinical trial, which may lead to operational execution challenges and increased complexity and burden for the sites and study participants.

Even though nearly everyone agreed on the advantages of a prospective evaluation because of the intrinsic greater reliability of controlled conditions, the majority concurred this is *not* essential, and hence it would be possible to assess treatment failure to past/current antidepressant treatment attempts, but only if properly documented *(as explained below)*, and in the past 2 years *(previous subsection, 3.2.2)*. We additionally recommend that these retrospective assessments must be based not only on standardised instruments to assess psychiatric history and previous treatments *(see section 3.5)*, but also on clinical documentation, such as pharmacy, hospital, or primary care records (which can also indicate some degree of adherence to the failed treatments). Such clinical documentation should be used also to carefully screen patients for previous episodes of mania, hypomania, or sub-threshold bipolarity, since it is desirable to have such population excluded, as discussed in 3.4.1.

Issues of debate

As for other recommendations in this report, we all recognised that the number of previous treatments required to define TRD and PRD is an arbitrary decision, but we also all agreed on the need of such decision in order to standardise recruitment of such samples in clinical studies and trials. An issue to discuss is that the precise threshold number for treatment failures may be influenced by the level of care. For example, in real-world clinical practice, it is more likely that a patient in primary care settings would be considered as resistant after one single treatment (first treatment attempt), while it is more common to consider the resistance after two (or three) treatments in secondary care (e.g., because the first treatment was administered in primary care). That is another reason why it is important to define a precise number of previous treatments to define TRD/PRD, allowing an across-the-board approach, applicable to all clinical scenarios.

There was more debate concerning the maximum number of failed treatments. The consensus (moderate) was that the number of failed trials should be used as a moderator in the analysis, rather than as an exclusion criterion. Some experts suggested to not include multiple resistant patients in TRD definition, in order to have a more homogeneous group. However, it is possible that some novel compounds might work even for patients with multiple failed treatments, and by excluding such subjects we would be negating the chances of testing new treatments for such individuals. This concept has been discussed by McAllister-Williams et al. ([31]) in their paper on MTR-MDD. The authors suggest that a threshold indicating MTR could be used in specific studies to look at the features of this subgroup of very resistant patients, or at least to describe how many such patients are present in a specific TRD cohort. However, it is important to consider that, with increasing numbers of failed treatments allowed, there could be increasing heterogeneity between patients in a study, although this is partly addressed by limiting the relevant period to the current episode and the last two years only.

Indeed, the issue of current/past episodes was one of the most debated. Even though we achieved only a weak consensus, there is a number of issues that needs to be taken into account when considering what defines the end of a previous episode and the beginning of the current one, and thus distinguish a ‘relapse’ (the return of MDD symptoms after remission, but before recovery) from a ‘recurrence’ (the onset of a new episode following recovery) ([50]). A patient is considered in remission if presenting no, or very few, symptoms (*as further discussed in section 3.5*), and recovered after a sufficient period – for example, of at least 4 months – of sustained remission ([19]). A patient experiencing a recurrence should be considered as having a new MDE, so distinct from the previous (past) episode. Nevertheless, importantly, relapses and recurrences occurring while taking an adequately dosed medication should be considered as the same (current) episode. Indeed, it is often problematic to correctly differentiate between true ‘treatment-resistance’ and other conditions, such as ‘tachyphylaxis’ (the loss of treatment efficacy after a transient improvement) ([51]), ‘pseudo-resistance’ (due to inappropriate or inadequate treatments, like suboptimal prescriptions) ([52]), or illness relapses and recurrences following successful response ([53]). Ideally, these conditions should be clearly distinguished from TRD/PRD, but more research is needed to accurately define them.

Strictly related to the issue of current vs. past episodes is the issue of what is the longest period of time within the current episode during which a valid retrospective assessment of the response can be

made. As mentioned above, only a weak consensus was reached around the 'two years' limit. Some experts suggested that only the past one year can be reliably assessed retrospectively, some suggested that up to 5 years should be assessed retrospectively, and some suggested to use a generic definition of 'only for a period when the assessor is convinced that adequate and clear evidence is available'. A number of apparent incongruities were highlighted based on any preferred definition. For example, with 5 years, you could include patients as TRD if they had been depressed for 5 years and took one medication 5 years ago for a few months with no response, and a second medication now also with no response. However, with the proposed 2 years, we would exclude patients if they had been on a single medication (and not responding) for the last two years, even if the current episode is longer than 2 years, as we would ignore all treatments prior to 2 years ago; this might be a problem, especially in countries like the UK, where 70% of patients who are on an antidepressant to which they are not responding have been on it for over a year, as we would be excluding a lot of patients who are TRD. As mentioned above, we can only recommend the consensus decision, in the hope that better research from more homogenous groups can provide answers to this kind of questions, and eventually lead to better definitions.

Recommendations

- Failing at least two prior antidepressant treatments – in the current episode – is necessary to define TRD.
    - *Level of consensus – Strong*

- A partial response to one single antidepressant treatment – in the current episode – is sufficient to define PRD.
    - *Level of consensus – Moderate*

- There is no maximum number of previous antidepressant treatments for patients recommended to be included in TRD (and PRD) regulatory trials, but all documentable life-time treatment response should be recorded to define the TRD (and PRD) stage of a patient (see also below).
    - *Level of consensus – Moderate*

- TRD (and PRD) definition (and differentiation) should be based on the current depressive episode only, and in the past two years only; if the current episode has lasted more than two years, treatments prior to the last two years should not be considered.
    - *Level of consensus – Weak*

- To define TRD (and PRD), response to previous antidepressant treatments (within the current episode and in the past two years) can be ascertained retrospectively using structured interviews and clinical documentation.
    - *Level of consensus – Moderate*

3.3. Type, dose, and duration of an 'adequate' antidepressant treatment

There is uncertainty in the definition of an adequate antidepressant treatment, regarding the type of antidepressant treatment (whether one with a primary indication for MDD only, or also other pharmacological agents), the role of non-pharmacological interventions, and the definition of adequate dose and duration ([21]). This section is predominantly dedicated to recommendations for the retrospective assessment of TRD/PRD of patients to be included in a clinical trial and to help the design of future clinical trials. It may be expected that with fast acting medications such as esketamine ([49]), and others in development, these definitions may need further revision.

3.3.1. Type of medications

Different classes and mechanisms of action of medications for MDD

Not all definitions of TRD specify whether the two failures could be with medications from the same class or whether they must be in different classes. A number of pharmacological classes of agents exist, based on their pharmacological mechanism of action. These include several tricyclic agents (TCAs), monoamine oxidase inhibitors (MAOIs), selective serotonin reuptake inhibitors (SSRIs), serotonin–norepinephrine reuptake inhibitors (SNRIs), and other atypical compounds such as the noradrenergic and serotonergic antagonists mianserin and mirtazapine. The EMA (2013) ([44]) specifically mentions that the two treatment failures could be with medications of "same or different class". Other guidelines, including the FDA guidance (2018) ([43]), do not make this distinction *(see also Table 1)*. The McAllister-Williams et al. ([23]) paper on DTD stresses the importance of adequate past treatments; however, they do not strictly define what should be considered as "usual treatment efforts" and include any pharmacological or non-pharmacological treatment modality. Moreover, they specify that treatment efforts "will depend on the health care setting and environment and relate to local treatment guidelines and practice". Although useful in clinical settings, this emphasis on clinical judgement and local environment limits the use of this approach for regulatory purposes.

To minimise variability associated with TRD/PRD definitions, we recommend that the two antidepressant treatment failures should consist of two medications of *different mechanisms of action*. The concept of 'different mechanisms' may overlap with 'different classes', although the pharmacological overlap between classes and mechanisms of action is not absolute; the Neuroscience-based Nomenclature (NbN) for psychotropic agents has greatly contributed to clarify this issue ([54]).

Finally, we consider two failed trials regardless of whether the two treatments are separated by a drug free period (and there is failure to both), are prescribed sequentially (switching, with failure to both), or whether one is prescribed as augmentation to the other (because of the first drug failing, and with the augmentation also failing).

Augmentation with medications not approved for the treatment of MDD

Following on from the point above, this report also recommends that both failures should be with two established medications for MDD, and not with, for example, a medication for MDD and an adjunctive drug (not approved for MDD) (either as switching or augmentation). The agreement on this point was

only moderate (and the issues of debate are extensively discussed below) but such a strict approach will improve the homogeneity of the TRD (and PRD) defined samples.

However, many patients who fulfil the proposed criteria for TRD may have failed to respond to both two (or more) medications *and* to one or more augmentation/combination strategies, including two medications (for MDD), or a medication (for MDD) and a different drug (not approved for MDD). Pharmacological augmentation interventions encompass mainly low-dose drugs for psychosis ([55]), but also mood stabilizers and other treatments, including NMDA-targeting medications, such as esketamine ([56]). We therefore discussed whether these patients should be considered to have TRD which is 'too severe' for inclusion in clinical trials. We already argued that there should *not* be an exclusion of patients from TRD studies because of a maximum number of failed medications (lifetime or in the last two years), but that this should be recorded and possibly used for subgroup analyses of MTR-MDD (*see Section 3.2*). Similarly, we recommend that patients who failed augmentation strategies should *not* be excluded from clinical trials on TRD/PRD (providing they fulfil the TRD/PRD criteria), but that this information should always be collected, for subgroup analyses. Of note, augmentation strategies are not classically proposed for TRD patients (only the fluoxetine plus olanzapine combination is licensed for TRD), but rather they are advised as a treatment strategy for PRD patients ([44]; [43]; [47]), even if this differentiation is not always straightforward *(see section 3.1)*.

Non-pharmacologic interventions

1. Brain stimulation

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, continuously (though reversibly) affecting neuronal functions ([57]). Because of the invasiveness of these intervention, patients selected for these treatments are often unique individuals who have failed an unusually large number of treatments, and thus may be different from other TRD patients in their characteristics. Based on these considerations, and only with moderate consensus *(see below)*, failure to respond to these two interventions is thus indicated as an exclusion criterion. However, failure to respond to other non-continuous/non-invasive brain stimulation interventions, such as electroconvulsive therapy (ECT) or as transcranial magnetic stimulation (TMS), is not considered an exclusion criterion (in the same way that failure to multiple medications and to augmentations strategies are also not considered exclusion criteria), but of course information on the past brain stimulation treatments should be collected for subgroup analyses and staging of the patients.

2. Psychotherapy

The experts highlighted that some patients may prefer to start their treatment with psychotherapy, and if they fail, they go on to medications, alone or in combination with the psychotherapy; thus, considerable debate ensued as to whether to include psychotherapy interventions as one of the two 'antidepressant' failed treatments required for the TRD definition. The majority view (albeit with moderate consensus) was that a failed course of psychotherapy should *not* be included as one of the two failed antidepressant treatments required for the definition of TRD, but this information should be always reported.

As for augmentation and brain stimulation, we also debated whether or not TRD patients who have also failed to respond to adequate psychotherapy trials are 'more' treatment-resistant and thus should be excluded. Again, as with the other strategies, and considering also the intrinsic difficulty in the standardized appraisal of previous psychological interventions, we recommend that failure to respond to a previous psychotherapy should *not* be an exclusion criterion for interventional studies, but that data should be properly collected and recorded, and possibly used for sub-analyses and staging.

3.3.2. Dosage

The majority of medications indicated for MDD have a wide therapeutic index, with a large range between the minimally effective dose and the maximum licensed dose. Moreover, some medications can be prescribed at higher than approved doses. There is limited and conflicting evidence for increased efficacy after medication dose increase, and most studies have found no benefit for dose escalation over staying on the minimum licensed dose, with an increased risk of side effects and discontinuation ([58]; [59]). However, meta-analyses of RCTs suggest that higher doses of medications used for the treatment of MDD may have superior efficacy compared with lower dosages ([60]; [61]) and that higher starting doses can be associated with higher rates of response ([62]).

After considerable debate, the consensus proposal is that the criteria of 'adequate dose' is *the minimal effective dosage*, that is, the minimal approved dosage.

3.3.3. Duration

After much debate, the most commonly held position considered four weeks to be the minimum duration of an antidepressant treatment before it is considered a failure. Importantly, this recommendation would not apply to ECT, which typically leads to a significant response within 2 weeks of initiating therapy, or to new fast acting medications, such as esketamine. Finally, in the context of regulatory clinical trials, patients' discontinuation of treatments may be considered as equivalent to a failure. However, we recommend that, because of the difficulty in distinguishing between non-response and intolerance when retrospectively assessing TRD/PRD, especially when people stop their therapeutic treatment before 4 weeks, patients' discontinuation of treatment *per se*, without clear evidence of lack of symptoms change, should not be considered as a demonstration of resistance.

Issues of debate

Type of medications

The rationale of choosing medications with different mechanisms of action was highly debated, and the overall consensus was moderate. Of course, this recommendation is applicable only to TRD, where two failed medications are recommended. This could seem rather restrictive compared with other guidelines stating that these two failed treatments could be with any two medications for MDD ([44]). As an example, the recent EMA approval of intranasal esketamine in TRD patients considered two treatment failures regardless from the mechanism of action ([44]). However, clinically there is some evidence that switching between medications of the same mechanism of action used under equivalent

regimens is less likely to produce a different clinical response, while switching between compounds from different classes could be more effective ([63]). The main issue of contention was how accurate our understanding of the concept of 'similar vs. different mechanism of action' is. While the NbN has improved our classification, it has not solved all the problems. Indeed, medications for MDD are usually pharmacologically complex, and their mechanism of action can be different from the biological (or neurobiological) effects or the downstream physiological effects. This report does accept that there is a certain degree of variability in judging whether, at an individual-level, two different failed medications could be considered 'from different classes' or not, based on the type and the dose of each medication, and their overall pharmacological profiles. Nevertheless, we also felt that allowing any sequential combinations of two (failed) medications for MDD regardless of the mechanism of action as criteria for TRD would include patients that have just taken potentially identical drugs, and could (relatively easily) respond to a different drug. Moreover, this definition (of two drugs from two different classes) is also consistent with the effectiveness of this approach from the clinical perspective of choosing the next-step treatment during clinical practice. As for other debated recommendations based only on some level of consensus, this report is keen to stress that some guidance is needed to provide the required evidence from future studies, in order to make more informed decisions in the future.

The issue of considering augmentation with medications not approved for the treatment of MDD in TRD/PRD definitions was also highly debated. Some experts supported the inclusion of patients who failed only one medication for MDD plus, for example, an adjunctive agent, like aripiprazole or quetiapine. This is also a valid point in relation to the debate about 'classes of antidepressants', as these drugs have several pharmacological targets and thus could theoretically be considered 'antidepressants of another class'. However, in the end the (moderate) agreement was to use a regulatory approach, that is, only considering augmentation with another approved medication. In fact, while studies indicate that adjunctive treatment with low doses of medications used in psychosis such as quetiapine ([64]), olanzapine ([65]), aripiprazole ([66]), brexpiprazole ([67]), risperidone ([68]), and ziprasidone ([69]), can be efficacious for TRD even in the absence of psychotic features ([70]), there are some geographically-relevant label limitations to their usage in TRD/PRD. Specifically, none of them are approved for clinical use in MDD, PRD, or TRD in the EU; and, in the US, the olanzapine/fluoxetine combination is approved for TRD, while aripiprazole and brexpiprazole are approved as adjunctive treatment for MDD (thus, also for PRD patients). Notably, quetiapine is approved in bipolar depression only (both in EU and US), and not for unipolar depression. Another medication with evidence of effectiveness in augmentation for MDD is lithium ([71]) but again this is not FDA or EMA approved in unipolar depression. Based on these considerations, and aware that definitions may diverge from the real-world clinical practice, the consensus was to require that both (failed) treatments should be with two approved antidepressant agents, again to limit variability within/between different TRD samples, and indeed to make sure no depressed patient would be defined TRD after treatment with only one proper antidepressant medication.

Results from the STAR*D study demonstrate that the chances of ameliorating depressive symptoms diminish as additional treatment strategies are used in switching or augmentation because of previous failure ([19]). This decrease is particularly relevant after level 2 of the study, in which different treatment options were provided besides the original SSRI, citalopram (given at level 1); this level randomly allowed for a switch (to bupropion, sertraline, or venlafaxine, or cognitive therapy) or augmentation

(citalopram plus bupropion, buspirone, or cognitive therapy). After the first two levels, almost half of the patients achieved remission (37% in level one and 31% in level 2). In level 3, patients were given lithium or a thyroid hormone (triiodothyronine) in augmentation to their reuptake inhibitor, or they were switched to mirtazapine or nortriptyline monotherapy; the proportion of patients achieving remission dropped to 14%. Finally, level 4 consisted of the non-selective and irreversible MAOI, tranylcypromine, or a combination of venlafaxine and mirtazapine ([19]), and at this level, the rate of remission was 13%. This evidence would imply that patients who have failed to 2 or more treatment strategies have progressively less chance to respond to switching to/adding a third one or even a fourth one, and this consideration would suggest the possibility that such patients should be excluded from TRD trials. However, notably, low-dose medications for psychosis were not used in STAR*D, which is a widely used strategy nowadays, and a failure to respond to some medications for MDD does not unequivocally imply failure to respond to new treatments.

There was some debate on the point of excluding patients who have failed DBS/VNS. Some experts argued that there is no evidence to justify the concept that there is a threshold that defines patients who are so treatment resistant that are never going to respond to anything – not even to medications or interventions that currently do not exist; and, indeed, this is the argument we have used in order to include TRD patients with failure to multiple medications and augmentation strategies in our definition. Other experts commented that it was arbitrary to include patients who failed to respond to ECT but not patients who failed to respond to VNS/DBS, although others replied that ECT, with its episodic administration and non-invasive approach, would still be offered to patients that have ‘less severe TRD’ compared with patients who are offered VNS/DBS, which are continuous and surgically invasive. Thus, there were serious concerns about including patients who failed to respond to VNS/DBS in a TRD definition that would also be used for regulatory clinical trials, as these patients usually have greater levels of resistance, and their inclusion would excessively increase the heterogeneity of the clinical phenotype, potentially distorting the results. Instead, we recommend these patients to be included in specific research studies and clinical trials for ‘refractory depression’.

Opinions from the experts were also quite discordant on the role of psychotherapy as one of the two antidepressant treatments for TRD definition. Psychotherapy has been proved to be an effective treatment for TRD patients ([72]; [73]). However, evidence is still sparce and psychological treatments are extremely variable in type and approach. Some argued that failure to respond to psychotherapeutic interventions should be regarded as equivalent to failure to respond to pharmacotherapy, if the interventions have been carried out as an evidence-based treatment for depression, such as cognitive behavioural therapy (CBT), interpersonal therapy (IPT), or cognitive behavioural analysis system of psychotherapy (CBASP). Others expressed the opposite opinion, i.e., against its inclusion in the definition, because of conflicting evidence of efficacy. Moreover, some argued that the failure to respond to psychotherapy does not necessarily imply a greater resistance to pharmacological treatments, while others claimed the opposite, namely, that patients who have failed to respond to psychotherapy as well as medication are ‘more’ treatment resistant. Others suggested that failure to respond to psychotherapy may instead presage a reduced likelihood to respond to placebo, and thus a better candidate for regulatory trials. In our recommendations, we decided *not* to include psychotherapy as one of the two failed treatments, even if we recommend including patients who failed psychotherapy and to always record it and analyse the data for subgroup analyses. Of note, clinical trials for purposes of regulatory approval, including platform trials, may in the future include

trials of psychotherapies or computer-based apps as new treatments or as comparisons. Even if this is still not available nowadays, we acknowledge that this recommendation will need to be revisited if trials of psychotherapies for depression will be conducted in parallel with medication trials.

Dosage

Opinions were particularly split regarding the recommended minimum antidepressant dosage to be used to define a failure. Of course, experts agreed that some patients only respond to higher doses, or the maximum licensed dosage, or also the maximum tolerated dosage (which can be even higher than the maximum licensed dosage). For example, venlafaxine is used at higher doses to induce a full noradrenergic effect ([74]), and there is some evidence that fluoxetine is more effective at a higher dose than the minimal licensed one ([75]). Thus, despite conflicting evidence, increasing the dosage can be a reasonable step in the clinical management of some patients, (especially in PRD); and indeed, it is established good clinical practice to increase the dosage when a medication is well tolerated and has produced a partial response at a low dosage. Nevertheless, there is no clear evidence in TRD *(see 3.3.2)*. Some experts highlighted the possible consequences of such an approach, for example that failure to respond to 50 mg of sertraline would be enough to count as one of the two required treatment failures. However, on balance, it was felt that, when applied to select TRD patients for clinical trials, using higher doses would restrict patients' selection excessively, to a point where many non-responders would be wrongly excluded from TRD trials because their doses were not pushed up to the maximum tolerated or licenced doses. Moreover, as discussed, the evidence for the efficacy of increasing dosage above the minimally effective one, especially for SSRIs, is scarce and too variable to include as an inclusion criterion for regulatory clinical trials. Importantly, pharmacogenetic interindividual differences in medication metabolism – fast or slow metaboliser status for different cytochromes – could also affect the effective or tolerated dosage and thus the response, but we have no systematic use of genetic information in TRD definitions *(see Future directions section, point 3.6.2)*.

Duration

In terms of minimum duration, we had experts suggesting both shorter and longer durations of treatment than the recommended four weeks. In particular, it was mentioned that just two weeks can be enough to observe an initial response, and therefore to predict non-response to treatment ([76]), and that exploring the trajectory of improvements, thus looking for early response, could be an important indicator of treatment resistance even following longer duration of treatments ([77]). In the end, the consensus reached was four weeks.

Besides minimum duration of treatments, some experts also raised the issue of the maximum duration. Indeed, patients who respond to treatment after 4 months or more (*see recovery definition in 3.3*; [78]) may also be considered as unresponsive, since this may be a spontaneous resolution of the current MDE rather than a response to the medication. They should therefore be distinguished from the cases where clinical improvement is experienced within 2-8 weeks after the treatment is initiated, which is more consistent with the typical time to improvement. While this report felt that there was not enough evidence to offer a clear indication, it was nevertheless felt that this point should be highlighted for future research.

Finally, we discussed that we should not consider the discontinuation of treatment (before the completion of the fourth week) for any cause, such as intolerance, as treatment resistance. Of course, these patients can still be included in TRD/PRD studies, provided they have two failures to at least 4 weeks of two medications for MDD, but not if they are only intolerant to two or more medications for MDD.

Recommendations

- To define TRD, the two different treatment failures must involve two established medications for MDD, with different mechanisms of action.
    - *Level of consensus – Moderate*
- Regulatory clinical trials for TRD (and PRD) may include patients who failed to respond (or partially responded) to augmentation/combination treatment strategies, but these need to be primarily based on medical records.
    - *Level of consensus – Strong*
- Regulatory clinical trials for TRD (and PRD) may include patients who failed to respond (or partially responded) to brain stimulation treatments, such as transcranial magnetic stimulation (TMS) and electroconvulsive therapy (ECT).
    - *Level of consensus – Moderate*
- Regulatory clinical trials for TRD (and PRD) should not include patients who failed to respond (or partially responded) to vagus nerve stimulation (VNS) and deep brain stimulation (DBS).
    - *Level of consensus – Moderate*
- Regulatory clinical trials for TRD (and PRD) may include patients who failed to respond (or partially responded) to structured psychotherapy.
    - *Level of consensus – Strong*
- The minimum effective dose of a medication indicated in MDD is enough to define a treatment failure for the purpose of establishing TRD (and PRD).
    - *Level of consensus – Moderate*
- For conventional medications indicated for MDD, a treatment given for at least four weeks in duration is sufficient to define a treatment failure for the purpose of establishing TRD (and PRD).
    - *Level of consensus – Moderate*
- Patients' discontinuation of treatment before the completion of the fourth week, should not be considered as a treatment failure for the purpose of establishing TRD (and PRD).
    - *Level of consensus – Strong*

3.4. Clinical presentation of TRD and PRD patients

As discussed in the *Introduction*, the diagnosis of MDD comprises a multitude of distinctive clinical features, with a multitude of signs and symptoms, and frequently with different psychiatric comorbidities. In this section, we will try to provide clear recommendations on how to deal with different clinical presentations in TRD/PRD research. While we believe the majority of the issues on this topic have consensus and are ready to be currently implemented (as summarised in the *Recommendations* at the end of this section), we will further discuss specific aspects to be addressed in future research in *section 3.6*.

3.4.1. TRD and PRD symptoms

It is important to examine the different specifiers of MDD (for example, with melancholic features, with atypical features, with psychotic features), as well as the presence of comorbid symptoms or conditions (for example, comorbid anxiety, bipolar depression). Potentially, some of these symptoms may be more difficult to treat compared with others or may respond differently to different types of treatments. As an example, MDD with anxious distress is associated with poor response to conventional medications ([29]; [79]; [13]; [80]); this specific subtype could therefore be considered as an indicator of higher treatment resistance compared with other forms of depression, as also shown by its better response with adjunctive treatment aripiprazole ([81]), quetiapine ([82]), and brexpiprazole ([83]). Similarly, comorbidity with personality disorders ([84]; [29]; [13]) or substance use disorders ([85]) *(see later in this section, point 3.4.3*) also makes depression more resistant to antidepressant treatment ([23]).

There was a very clear consensus to consider all specifiers of depression (melancholic, atypical, anxious, psychotic, mixed) within the TRD/PRD definition, *except for bipolar depression*, as this is a separate illness, part of bipolar disorder. Moreover, the majority view was that, at present, the assessment of changes in severity (upon which the TRD/PRD definitions rely) should include the entirety of symptoms that a patient displays, using broad-spectrum symptoms scales or checklists (*see below, 3.5*). However, sub-analyses on specific clusters of classic symptoms (such as suicidal, atypical, or psychotic symptoms) could be used to generate hypotheses for future studies on targeted medication licenses and for clinical trials using symptoms-based subgroups of patients as inclusion criteria. This may be especially important for future research directions, because some symptom profiles might be best treated by targeting relevant biological mechanisms, such as inflammation for anhedonia and vegetative symptoms, or hypothalamic-pituitary-adrenal (HPA) axis overactivity for psychotic symptoms – an area of research within the regulatory framework that is still currently at its infancy *(see Future Research below)*.

3.4.2. Comorbidities

As mentioned above, comorbidity with personality disorders or with substance use disorders can exacerbate a depressive disorder, or make antidepressant treatments less effective ([86]; [87]). The inclusion of these patients in regulatory clinical trials could thus influence the results. For example, patients with a primary diagnosis of personality disorder (especially borderline personality disorder) may frequently meet criteria for MDD, but they are often unlikely to respond well to antidepressant treatments and may hence mimic a non-response. Also, active substance users could distort the results of clinical trials because of pharmacological interactions (both pharmacokinetic and

pharmacodynamic), side effects, and mood symptoms during substance abuse or withdrawal which may appear as a 'phenocopy' of MDD. However, it is important to also point out that these conditions are so frequently co-morbid with depression that they cannot be routinely excluded, otherwise there is a risk that the proposed TRD/PRD definitions are so strict that the results are not generalisable to patients seen in everyday clinics (as also discussed in the *Methods*).

The FDA guidance for inclusion in TRD/PRD studies 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)" ([43]). Indeed, the FDA does not explicitly take position on the inclusion or exclusion of personality disorders, and it explicitly *encourages* to consider patients 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 focus of the guideline ([44]).

Our conclusion was that these conditions must be excluded, but only if they reflect a *primary* diagnosis of a personality disorder or substance use disorder, that is, diagnosed before, and independently from, MDD. Moreover, in accordance with the FDA draft guidance ([43]), we recommend excluding subjects with an *active* substance use disorder, intended as a severe substance use disorder and not in remission (according to DSM-5 criteria), yet allowing for the inclusion of subjects with a history of substance abuse or with mild or moderate substance abuse with onset subsequent to the MDD diagnosis.

Similarly, this approach is also advocated for other psychiatric comorbidities, such as generalized anxiety disorder (GAD), obsessive–compulsive disorder (OCD), attention deficit hyperactivity disorder (ADHD), and post-traumatic stress disorder (PTSD): to exclude these conditions if they are the primary diagnosis, but allow the inclusion of these patients if the conditions develop after the MDD. Also, somatic comorbidities should be systematically recorded, but not excluded *a priori*. This should apply also to conditions such as inflammatory, neuroendocrine, and metabolic diseases, which importantly can influence the response to treatments ([88]; [89]; [90]).

Issues of debate

The main issue with psychiatric comorbidities is that their onset, independently from MDD, should be corroborated, when possible, by clinical records or other collateral history, and not only by patients' accounts. Moreover, because of the symptoms overlap, it is sometimes difficult to distinguish diagnostically between MDD and other psychiatric conditions, such as GAD ([91]). In the future, this issue could be addressed by a more dimensional (and transdiagnostic) approach *(see also section 3.6)*. The experts also suggested that individuals with a primary diagnosis of personality disorder or with active and severe substance use (thus, excluded from the present definition) should be included in future research, on treatments specifically targeting MDD in those subjects.

A related issue is how to weigh differences in sociodemographic characteristics, particularly age. Indeed, depression in elderly may be associated with delays in antidepressant response and greater susceptibility to side effects ([92]). Moreover, the prevalence of MDD increases in late life, and it is often comorbid with other physical illnesses and consequent polymedication ([93]). It is also crucial to properly assess MDD symptoms in older patients and to distinguish early-onset depression from late-onset depression, as this latter may be subsequent to organic diseases, such as prodromal states of dementia ([94]). Because of these differences and complexities, most of the clinical trials for regulatory purposes set a maximum age for inclusion (usually 60-65 years). Given the aim of this document and to avoid excessive heterogeneity *(section 2.4, General limitations of the report)*, we aligned with this view. All the recommendations made throughout this report must be considered for adults aged between 18 and 65 years. Instead, we recommend including patients aged over 65 years in different and specific research studies and clinical trials, as seen for patients who failed to respond to VNS and DBS *(section 3.3)*.

Recommendations

- No specific types of symptoms of MDD should either be prerequisite, or excluded, from the definition of TRD (and PRD), but symptoms and specifiers should always be recorded.
    - *Level of consensus – Strong*

- Patients with bipolar depression should be excluded from TRD (and PRD) studies, as this is a separate condition from unipolar depression (MDD).
    - *Level of consensus – Strong*

- Patients with comorbid personality disorders or other mental disorders should be excluded from TRD (and PRD) studies only when their onset is properly documented as independent and antecedent to the MDD diagnosis.
    - *Level of consensus – Moderate*

- Patients with comorbid substance use disorder that is active and severe should always be excluded from TRD (and PRD) studies, independently from the onset; in contrast, patients with comorbid substance use disorder that is active and mild/moderate should be excluded from TRD (and PRD) studies only when the onset is properly documented as independent and antecedent to the MDD diagnosis.
    - *Level of consensus - Moderate*

3.5. Diagnostic tools and measures of outcome

The experts were keen to establish a consensus on the best psychometric tools to measure antidepressant response (and thus TRD diagnosis and staging) both retrospectively, to improve the way we diagnose TRD (and PRD) before entering a trial, but also prospectively, as tools to use in future regulatory trials for these indications.

3.5.1. Historical assessment of treatment resistance and diagnosis of TRD/PRD

The minimum required approach to assess treatment resistance consists of using structured clinical interviews for the diagnosis of MDD, such as the Structured Clinical Interview for DSM (SCID) ([95]; [96]) and the Mini-International Neuropsychiatric Interview (MINI) ([97]), together with specific scales to assess the patient's antidepressant history, such as the ATRQ (Massachusetts General Hospital Antidepressant Treatment Response Questionnaire) ([98]), and the Antidepressant Treatment History Form (ATHF) ([99]). The essential value of the structured interviews is to confirm the diagnosis of MDD and to assess the presence of TRD throughout the patient's entire clinical history, also allowing the identification of comorbidities. The antidepressant history scales, ATRQ and ATHF, examine the previous antidepressant treatments as told by the patient, and both have pros and cons: the ATRQ is easier to use for both patients and clinicians, but includes only the patients' recollection of how much they improved on any antidepressant trial ([100]); the ATHF has the advantage of integrating clinical judgement in evaluating the adequacy of a treatment, but is quite long, and thus has been recently revised and updated in a short form (ATHF-SF) ([101]).

A highly more structured method is to use staging models ([102]; [22]). Probably the most commonly used model is the Thase and Rush method ([103]), with TRD (as described in this report) being equivalent to stage 2 of this model, which considers the "failure of at least 2 adequate trials of medications (of different classes)". Despite its wide use, the Thase and Rush approach has some limitations, including the absence of a clear definition of what is an appropriate antidepressant trial in terms of dosing and duration, with the risk of also including interventions which did not have the opportunity to show an effect because of either too low of a dosage or too short of a duration, or both. Furthermore, MDD patients do not necessarily progress in a linear way through the proposed hierarchy of medications (SSRIs < TCAs < MAOIs), so more treatment-resistant patients may not necessarily reach higher stages of their scheme. Finally, there is no consideration of augmentation strategies and psychotherapy.

Another staging model is the Massachusetts General Hospital Staging model (MGH-s) ([24]). This model defines TRD as the "failure to respond to at least one antidepressant trial of standard doses lasting 6 weeks or more", so it is less restrictive than the EMA and FDA criteria as well as the consensus recommendation proposed in this report, albeit more restrictive in duration. This model generates a continuous score, with different points attributed to different clinical features, reflecting the total level of resistance. Interestingly, this scale includes "optimization of dose and duration, augmentation or combination options", as additional steps beyond the adequate trial, attributing the same score to each of these steps.

A third, more recent tool, is the Maudsley Staging Model (MSM) ([104]). As with the MGH-s, the MSM also considers one antidepressant treatment failure as enough to define TRD; however, it supports the notion that TRD exists as a continuum, further elaborating that "failure of the first treatment is influential in treatment resistance and may be a useful starting point in any measure of this conceptual continuum" ([22]). The MSM has been proven to have a very high predictive utility for future treatment response ([102]; [22]); it can be used both as a continuous score (from 3 to 15) and divided into three ordinal categories (mild: 3-6, moderate: 7-10, severe: 11-15). The MSM also allows a 'dimensional' staging, based on duration, severity, and treatment, and it is also predictive in non-antidepressant-treated patients, being a more general indicator of duration, chronicity, and severity, beyond treatment

response alone ([105]). Finally, the MSM also allows the assessment of previous antidepressant treatment failure, using the Maudsley Treatment Inventory (MTI), which is a new tool to assess a patient's antidepressant treatment history developed for the purposes of completing the MSM ([32]). As a potential limitation, MSM arbitrarily divides the duration of illness into three categories (acute: ≤ 12 months, sub-acute: 12-24 months, chronic: > 24 months), giving a 'higher score' for TRD for longer duration, irrespective of the treatment history, while one could argue that longer duration of illness is not necessarily an index of treatment-resistance.

Among the experts, there was a substantial agreement that staging models are preferred compared to clinical interviews for the diagnosis of TRD/PRD, even if many experts recognize the validity of the ATRQ to assess patients' antidepressant treatment history, often accepted in regulatory clinical trials. The largest consensus emerged in favouring the MSM, mostly because it captures the dimensional nature of TRD (although some debate emerged on this point, *see below*). However, both the Thase and Rush and the MGH-s models were considered valid alternatives.

3.5.2. Assessment of depressive symptoms and response to antidepressant treatment

The presence and severity of depression are usually established through depression assessment scales: clinician-administered scales include mainly the Hamilton Depression Rating Scale (HAMD 17, 21, 24 items) ([106]), the Montgomery-Asberg Depression Rating Scale (MADRS10) ([107]), and the Quick Inventory of Depressive Symptomatology (QIDS) Clinician Rating (QIDS-C) ([108]), while self-reported instruments include mainly the Beck Depression Inventory (BDI) ([109]), the Patient Health Questionnaire-9 Item (PHQ-9) ([110]) and the QIDS Self-Report (QIDS-SR) ([108]). When possible, these scales should be administered before and after a specific treatment has started, to measure the variation in depressive symptoms, but they can also be used as a single measure to determine the current presence and severity of the depression (for example, in people taking a medication), even if no pre-treatment measures are available. Outcome measures are crucially important because they are the most reliable method to define (either for the current episode, or prospectively) TRD/PRD patients, and to distinguish them from full responders.

In addition to these depression-specific scales, there are also more generic instruments to assess psychiatric global status which are frequently used, such as the Clinical Global Impression (CGI) scales ([111]). Furthermore, even antidepressant treatment responders may continue to have significant residual symptoms and functional impairment. A valid instrument frequently used to assess functional impairment is the Sheehan Disability Scale (SDS) ([112]). Finally, patients often define treatment success quite differently, with a heavy emphasis on broadly-based, functional outcomes rather than remission of individual symptoms. Their perspective could also be included in the outcome measures through patient-reported outcomes (PROs), such as quality of life, individual's perspective, wellbeing, impact of symptoms, and social and occupational functioning ([113]).

The ideal treatment goal of every treatment is remission, considered as the absence of a relevant MDD symptomatology, as we discussed in *3.2 Issues of debate*. Specifically, a patient is considered as a remitter when the score in a specific scale is below a fixed cut-off value, usually – but not universally – considered as being ≤ 7 at the HAMD17 ([114]) or ≤ 10 at the MADRS10 ([115]; [116]) or equivalent. Given that 'remission' is a static construct, as it does not consider the variation in

symptoms over time, it can be used at a single time-point assessment. However, the most widely used approach to assess the outcome of an intervention is 'response' (≥50% complete response, 25-49% partial response (used for PRD), <25% non-response (used for TRD), *as discussed in section 3.1*).

Of note, as part of this consensus process, we also discussed if these criteria of remission and response were valid in the context of TRD/PRD, or needed to be changed, for example redefining response criteria in a less conservative way, so to identify more cases which are responding to a treatment but would still be considered non-responders using the 'usual' thresholds. Another approach in order to optimize our chances of detecting an improvement in TRD/PRD patients in the course of clinical trials may be the use of shorter versions of the most commonly used scales, such as the HAMD6 and the MADRS6, which are focused only on core symptoms of depression, not measuring accessory symptoms or 'trait' features; the latter scales tend to be more sensitive to changes in severity, and thus more suitable to measure treatment response (although at the cost of missing the entirety of the clinical picture).

The experts agree that the favoured instrument was the clinician-administered MADRS10, rather than the HAMD17, because the latter puts too much weight on anxiety and somatic symptoms, and has a lesser focus on core depressive symptoms. Moreover, most of the experts believed that the association of a clinician-administered scale with a patient-reported instrument, especially the QIDS-SR, is the most helpful in capturing the whole clinical picture ([117]) *(see also point 3.6.3)*. Also, there are suggestions that a mismatch between scores in clinician-administered and self-reported scales is a poor prognostic sign ([118]), supporting the position endorsed in this report. Finally, the majority of experts agreed that we do not need to relax criteria for response or remission, or to switch to shorter versions of the scales.

Issues of debate

There are two elements that we should clarify when speaking about historical assessment of treatment resistance. One is that it is critical to have tools to aid the systematic collection of data. Secondly, it is important to be able to give a score for degree of TRD/PRD. Staging models, as debated, are currently the most valid instrument to be used, but it is not necessarily the only possible one. This 'score' could be a 'stage', but also a 'scoring system' or a 'semi-structured tool'. Thus, and many contributors agreed on this point, a more flexible continuous scale may be more useful. It is also important to highlight that a raw staging score *per se* (like in the MSM) does not indicate the different elements of the clinical history, like treatments failed or illness severity and duration, which should all be separated out when reporting it, so that the reader can understand the patient population being studied. Future directions of research to address the limitation of the available tools will be discussed later in *3.6*.

Concerning the assessment of depressive symptoms, it is important to highlight that currently available outcome measurement tools have some intrinsic problems. Firstly, there is a lack of content overlap between the different tools ([119]). Therefore, measuring different symptoms, they could potentially provide different classifications of MDD patients. In addition, the measurement error at each assessment and the lack of measurement invariance across timepoints, may contribute to

complicate the identification a population based on a percentage reduction on clinical scales. Some contributors also pointed out a potential limitation of using the MADRS, which as discussed, only measures core depressive symptoms, and not, for example, anxiety symptoms. While this is generally an advantage when evaluating MDD scores, it may not catch the clinical complexity of the patient, thus other specific scales could be used to assess different MDD specifiers (e.g., Hamilton Anxiety Rating Scale (HAM-A) ([120]), for anxiety symptoms). 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 be a trigger for relapse. We therefore recommend, where possible, to perform the broadest assessment possible with full-scales or additional scales, in order to capture the entirety of symptoms and identify potential treatments efficacious on specific clusters of symptoms *(see also Future directions, 3.6.1)*.

Finally, we acknowledge that, especially in research settings, the capacity to measure response occurring within hours rather than days is becoming increasingly relevant. This is clearly the case with esketamine, but also, for example, some forms of DBS and psilocybin ([121]). Although, as discussed, this report does not recommend the routine use of shorter versions of traditional MDD scales, we acknowledge these may more adequately cover the symptoms of relevance for patients in these specific cases. In fact, traditional scales are not the most effective to be used for this purpose as they include items such as sleep and appetite, which clearly cannot vary in such a brief time. Instruments that are sensitive to measure short term changes should be developed and validated for future research.

Recommendations

- Maudsley Staging Model is the suggested instrument to define the degree of treatment resistance historically.
    - *Level of consensus - Moderate*

- Clinician administered MADRS10 is the suggested outcome instrument to assess treatment response (and remission) and, together with patient-reported QIDS-SR, can be used to assess TRD and PRD status.
    - *Level of consensus – Moderate*

- Criteria for remission, response, and partial response should not be relaxed in regulatory clinical trials for TRD (and PRD); shorter versions of the traditional scales, such as the HAMD6 and the MADRS6, should not be currently preferred to traditional scales, although may become more relevant in the future with fast-acting interventions.
    - *Level of consensus - Moderate*

3.6. Future directions

In this section, as mentioned in the *Methodology (point 2)*, we discuss the potential implications for future TRD and PRD research, which include some of the most hotly debated issues among the experts. This section makes recommendations that are not immediately applicable to clinical trials but

should be addressed by collecting new data as part of clinical research, longitudinal cohorts, or electronic health records.

3.6.1. Clinical phenotypes of TRD and PRD and dimensional approach

As mentioned in the *Introduction* and in the *Clinical presentation* section, it is essential for MDD research to move away from the idea that all depressed patients are the same. Even though it is compellingly clear from the clinical picture, and response to treatments, that MDD is *not* a single entity, this is the view commonly endorsed by clinicians, pharmaceutical companies, and regulatory agencies. In contrast, targeting treatments to 'biologically-based subgroups of patients' can bring 'personalised medicine' to psychiatry, akin to what has happened in oncology where cancer types are treated more in relation to their biological and molecular features than to their diagnostic classification. In MDD, genetic factors are far more complex than in cancer, with a myriad of risk genes rather than few genes of relatively large effect. However, we could be focusing on MDD patient subgroups based on specific clusters of symptoms, especially if mediated by a specific biological abnormality. This concept is also valid from a regulatory perspective and for the development of trial platforms; future trials may aim to enrich for symptom domains based on the medication to be tested. In order to do this, we need research using full scales assessing a broad range of depressive symptoms (such as the HAMD28), or scales assessing specific symptoms that may have an underlying biological mechanism, such as anhedonia, putatively linked to inflammation and dopamine deficit ([122]). These would also allow research in specific symptoms that remain *after treatment,* and thus in the development of adjunctive therapies targeting residual depressive symptoms. In addition, new digital applications for monitoring patients' symptoms, mainly as self-assessment instruments, may be implemented for future research.

Another approach to be examined in future research comes with the use of a 'continuum' approach in defining treatment-resistance. For example, in hypertension there is a clear definition for treatment-resistant hypertension, that is, resistance to maximal doses of 3 antihypertensives of different classes. This is different from other categories but on a continuum of severity of resistance which progresses to controlled resistant hypertension (blood pressure controlled on 4 or more antihypertensives) and then to refractory hypertension (blood pressure uncontrolled despite maximal doses of 5 or more antihypertensives of different classes) ([123]). Interestingly, in treatment-resistant hypertension the therapeutic approach is multidimensional, with a strong emphasis not only on clinical data (symptoms and treatments), but also on risk factors, demographic data, and lifestyle factor, mainly diet and physical exercise – something which has not been tested coherently in TRD yet, but could be extremely important in the future.

Another approach to be examined in future research comes the use of a 'continuum' approach in defining treatment-resistance. For example, in hypertension there is a clear definition for treatment-resistant hypertension, that is, resistance to maximal doses of 3 antihypertensives of different classes. This is different from other categories but on a continuum of severity of resistance which progresses to controlled resistant hypertension (blood pressure controlled on 4 or more antihypertensives) and then to refractory hypertension (blood pression uncontrolled despite maximal doses of 5 or more antihypertensives of different classes) ([123]). Interestingly, in treatment-resistant hypertension the

therapeutic approach is multidimensional, with a strong emphasis not only on clinical data (symptoms and treatments), but also on risk factors, demographic data, and lifestyle factor, mainly diet and physical exercise – something which has not been tested coherently in TRD yet, but could be extremely important in the future.

3.6.2. Genetics and biological markers

All the experts agreed that biomarkers are extremely important for future research, but a consensus emerged that there are no genetic/biological markers currently ready to use for inclusion/exclusion of TRD-PRD patients in clinical trials. However, several biomarkers held promise for future research, especially as part of more complex diagnostic and predictive algorithms, which include clinical, cognitive, blood based, genetic, and neuroimaging markers.

Among the most discussed genetic markers of treatment response in MDD, the experts highlighted cytochrome P450 polymorphisms, especially of the enzymes CYP2D6 and CYP2C19 (Kirchheiner *et al.*, 2004), which could affect an individual's metabolism of different compounds. Also, biomarkers related to immune function were discussed as markers of TRD, including gene expression signatures ([124]; [125]), cellular immunophenotype ([126]), and soluble factors such as C-reactive protein (CRP) ([127]; [128]). Other candidate markers include genetic profile or expression levels of biomarkers of HPA axis activity ([129]; [130]); channels controlling efflux of drugs from brain, e.g., ABCB-1 ([131]); serotonin transporter promoter ([132]); serotonin 1A or 2A receptors ([133]; [134]); olfactomedin-4 ([135]); and brain-derived neurotrophic factor (BDNF) gene ([134]).

Even though we recognize the validity of these encouraging findings, we support the need for further research before the inclusion of specific biomarkers in recommendations for regulatory trials. With this in mind, of course we advocate that current and future clinical trials at least collect the biological samples needed for subsequent testing of candidate biomarkers (DNA, whole blood mRNA, serum or plasma), or candidate biomarkers that are clearly related to the drug's mechanisms of action (for example, an immune biomarker for an anti-inflammatory drug), so that post-hoc analyses on sub/stratified samples are possible, and can inform future confirmatory trials. Interestingly, the development of biomarkers for subtypes of depression could be used to assess outcome for Phase II/Proof of Concept studies that have the potential to improve our ability to eliminate treatment targets that are not ultimately going to pan out earlier (at lower cost) and more reliably than currently possible; this approach is referred to as 'Fast-Fail' ([136]).

3.6.3. Patients' preferences and perspectives

There was a substantial agreement on the value of including patients' preferences when evaluating treatment resistance. This is another important point which could help to move MDD research away from a one-size-fits-all approach. Indeed, there was a general consensus on the importance of looking at TRD and PRD from a patient's perspective, thus not only including patients' preferences, but also integrating them into the definition, together with their subjective experience, to better understand this condition.

To develop this report, we involved patients' representatives in the stakeholders meeting and feedback, to fully consider their viewpoint (*see Annex V: Full list of contributors to D4.1 and acknowledgement*). An important message that came out of people with lived experience is that clinical scales often do not adequately consider the patients' perspective, with a consequent potential discrepancy between clinician- and patient- reported outcomes (PROs) ([137]). The key example of this is the fact that a considerable proportion of remitted depressed patients (remitted according to the traditional clinical scales) do not consider themselves as remitted, because of the persistence of specific symptoms not adequately recorded, such as persistent residual cognitive symptoms or poor functionality, which obviously affect the outcome ([138]).

As mentioned above, this discussion is relevant for the development of future protocols to assess depressed patients longitudinally and to improve the future clinical trials for TRD/PRD, both of which are key aims of this IMI EU-PEARL programme. Indeed, various approaches have been suggested to include patients' preferences and perspectives in TRD/PRD definitions, including the importance of recording which symptoms are the most disabling for the patient and what are the relevant functional outcomes, which may be more important than purely symptomatic ones. Some experts suggested the use of specific scales, such as quality of life (QoL) scales, specific checklists, or a visual analogue scale (VAS). Because there is no evidence of clear superiority of a single instrument over others, we do not recommend the use of any specific scale. It was also discussed whether PROs should be included, as opposed to being reported as a secondary outcome measure in trials, in the definition of TRD and PRD.

The recommendation is to integrate patients' preferences and perspectives in future studies, as also clearly highlighted by stakeholders. While we endorse the use of self-reported measures of depression, such as QIDS-SR, the feedback from the stakeholder clearly indicated that these instruments, whilst offering a patient's viewpoint, still do not provide a comprehensive assessment of the patient's preferences, perspectives, and reported outcomes, which should be included in a standardized manner in future protocols.

Ultimately, future research should better identify the outcomes which are important to patients and consider them to better define what clinical measures to prioritise. For example, it is important to understand the patient target, whether it is a return to premorbid status or reaching an 'optimal' functioning level. Setting the bar too high could be counterproductive for the patient himself, potentially masking an effective treatment.

3.6.4. Adherence

Finally, 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 restricted time periods ([139]). However, MDD individuals have typically high reported rates of non-adherence (up to more than 50%) ([140]; [141]). It is therefore vital to properly confirm the patient'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. Several factors have been associated with future poor adherence to antidepressant treatments, including personal and cultural attitudes, such as stigma and health beliefs; sociodemographic features, such as unemployment and living alone; and clinical features, such as

comorbid personality disorders and alcohol/substance abuse ([142]). The assessment of treatment adherence can be quite 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 patients' plasma (which would also allow the recognition of fast and slow metabolisers), even though it may increase patients' 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.

Experts also suggested the potential use of other specific methods to assess compliance in clinical trials. These could, for example, detect specific markers, such as odorous compounds or olfactory markers, upon breath exhalation after medication is taken (Xhale®), or use of digital applications, such as diaries and reminders, or artificial intelligence to confirm the identities of the patient and the medication and verify intake (AiCure®). Finally, future research should assess whether it may be useful to measure compliance in the run-in period, prior to the clinical trial is initiated, to screen out non-compliant patients ([143]). This could be integrated into future protocols for regulatory purposes, to guarantee the reliability of findings. It is however important to note that the FDA does not accept an analysis that excludes patients not compliant to the previous treatment based on a blood dosing of the medicine. This created a debate. For some contributors, indeed, ensuring minimal compliance, estimating plasma penetration, and verifying adequacy of the dispensing, is absolutely essential. Therefore, in their opinion, assessing participant's adherence, by at least a plasma sampling at estimated steady-state and at the end of the study should be mandatory. Our position is that, at the current state, such a recommendation for current implementation is not possible. However, we recommend assessing the usefulness of different methods to measure a patient's adherence for a potential (and desirable) future implementation.

Recommendations

- Future research should be more patient-centred, recognizing, and targeting different clinical phenotypes of TRD and PRD underpinned by a specific biological mechanism.
    - *Level of consensus - Strong*

- 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 patient.
    - *Level of consensus – Strong*

- Currently, no biomarker has been validated in clinical practice or in clinical trials to identify TRD (and PRD) patients, or to further stratify them; however, collection of biological samples for subsequent subgroup or stratified analyses is recommended.
    - *Level of consensus - Moderate*

- Patients' preferences, perspectives, and reported outcomes should be included in future TRD (and PRD) diagnostic tools and outcome measures.

- *Level of consensus – Strong*

- 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.
    - *Level of consensus - Moderate*

4. Conclusion

There is a highly recognised need for clear and consistent definitions of treatment response in MDD for regulatory clinical trials, which can no longer be further delayed. This consensus document aims to fill the gaps in knowledge within TRD (and PRD) research, providing straightforward and replicable criteria agreed among a large group of experts, including clinicians, academicians, researchers, members of industry, regulatory agencies, and patients' representatives.

Reaching consensus of key definitions is the first deliverable of the WP4 EU-PEARL project, the larger goal of which is to design a Phase II/Proof of Concept platform trial. Standardised research is the only way to advance the MDD field towards tailored treatments. This 'precision medicine' would finally help to deliver better care for patients suffering from this severely disabling illness, which remains too often not adequately treated.

5. References

1. American Psychiatric Association. *Diagnostic and Statistical Manual DSM 5*.; 2013. doi:10.1176/appi.books.9780890425596.744053
2. World Health Organisation. ICD-11 for Mortality and Morbidity Statistics (ICD-11 MMS) 2018 version. *Https://IcdWhoInt/Browse11/L-M/En*. 2018.
3. Otte C, Gold SM, Penninx BW, et al. Nature disease primer: Major Depressive Disorder. *Nat Publ Gr*. 2016. doi:10.1038/nrdp.2016.65
4. James SL, Abate D, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018. doi:10.1016/S0140-6736(18)32279-7
5. Mueller TI, Leon AC, Keller MB, et al. Recurrence after recovery from major depressive disorder during 15 years of observational follow-up. *Am J Psychiatry*. 1999. doi:10.1176/ajp.156.7.1000
6. Solomon DA, Keller MB, Leon AC, et al. Multiple recurrences of major depressive disorder. *Am J Psychiatry*. 2000. doi:10.1176/appi.ajp.157.2.229
7. Ishak WW, Mirocha J, James D, et al. Quality of life in major depressive disorder before/after multiple steps of treatment and one-year follow-up. *Acta Psychiatr Scand*. 2015. doi:10.1111/acps.12301
8. J.-P. L, M. B. The increasing burden of depression. *Neuropsychiatr Dis Treat*. 2011.
9. Mols F, Husson O, Roukema JA, van de Poll-Franse L V. Depressive symptoms are a risk factor for all-cause mortality: Results from a prospective population-based study among 3,080 cancer survivors from the PROFILES registry. *J Cancer Surviv*. 2013. doi:10.1007/s11764-013-0286-6
10. Plana-Ripoll O, Pedersen CB, Agerbo E, et al. A comprehensive analysis of mortality-related health metrics associated with mental disorders: a nationwide, register-based cohort study. *Lancet*. 2019. doi:10.1016/S0140-6736(19)32316-5
11. Momen NC, Plana-Ripoll O, Agerbo E, et al. Association between Mental Disorders and Subsequent Medical Conditions. *N Engl J Med*. 2020. doi:10.1056/nejmoa1915784
12. Hasin DS, Sarvet AL, Meyers JL, et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry*. 2018. doi:10.1001/jamapsychiatry.2017.4602
13. Bartova L, Dold M, Kautzky 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. doi:10.1080/15622975.2019.1635270
14. Sharma V, Khan M, Smith A. A closer look at treatment resistant depression: Is it due to a bipolar diathesis? *J Affect Disord*. 2005. doi:10.1016/j.jad.2004.01.015
15. Stahl SM, Morrissette DA, Faedda G, et al. Guidelines for the recognition and management of mixed depression. *CNS Spectr*. 2017. doi:10.1017/S1092852917000165

16. Maj M, Stein DJ, Parker G, et al. The clinical characterization of the adult patient with depression aimed at personalization of management. *World Psychiatry*. 2020. doi:10.1002/wps.20771

17. Arnow BA, Blasey C, Williams LM, et al. Depression subtypes in predicting antidepressant response: A report from the iSPOT-D trial. *Am J Psychiatry*. 2015. doi:10.1176/appi.ajp.2015.14020181

18. Cipriani A, Furukawa TA, Salanti G, 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. doi:10.1016/S0140-6736(17)32802-7

19. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. *Am J Psychiatry*. 2006. doi:10.1176/ajp.2006.163.11.1905

20. McIntyre RS, Filteau MJ, Martin L, et al. Treatment-resistant depression: Definitions, review of the evidence, and algorithmic approach. *J Affect Disord*. 2014. doi:10.1016/j.jad.2013.10.043

21. Gaynes BN, Lux L, Gartlehner G, et al. Defining treatment-resistant depression. *Depress Anxiety*. 2019. doi:10.1002/da.22968

22. Salloum NC, Papakostas GI. Staging treatment intensity and defining resistant depression: Historical overview and future directions. *J Clin Psychiatry*. 2019. doi:10.4088/JCP.18r12250

23. McAllister-Williams RH, Arango C, Blier P, et al. The identification, assessment and management of difficult-to-treat depression: An international consensus statement. *J Affect Disord*. 2020. doi:10.1016/j.jad.2020.02.023

24. Fava M. Diagnosis and definition of treatment-resistant depression. *Biol Psychiatry*. 2003. doi:10.1016/S0006-3223(03)00231-2

25. Kessler RC, Berglund P, Demler O, et al. The Epidemiology of Major Depressive Disorder: Results from the National Comorbidity Survey Replication (NCS-R). *J Am Med Assoc*. 2003. doi:10.1001/jama.289.23.3095

26. Berlim MT, Turecki G. Definition, assessment, and staging of treatment-resistant refractory major depression: A review of current concepts and methods. *Can J Psychiatry*. 2007. doi:10.1177/070674370705200108

27. Sheehan D V., 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. doi:10.1097/YIC.0b013e328341bb5f

28. Bennabi D, Aouizerate B, El-Hage W, et al. Risk factors for treatment resistance in unipolar depression: A systematic review. *J Affect Disord*. 2015. doi:10.1016/j.jad.2014.09.020

29. Souery D, Oswald P, Massat I, et al. Clinical factors associated with treatment resistance in major depressive disorder: Results from a European multicenter study. *J Clin Psychiatry*. 2007. doi:10.4088/jcp.v68n0713

30. Dudek D, Rybakowski JK, Siwek M, et al. Risk factors of treatment resistance in major

depression: Association with bipolarity. *J Affect Disord*. 2010. doi:10.1016/j.jad.2010.03.001

31. McAllister-Williams RH, Christmas DMB, Cleare AJ, et al. Multiple-therapy-resistant major depressive disorder: A clinically important concept. *Br J Psychiatry*. 2018. doi:10.1192/bjp.2017.33

32. Fekadu A, Donocik JG, Cleare AJ. Standardisation framework for the Maudsley staging method for treatment resistance in depression. *BMC Psychiatry*. 2018. doi:10.1186/s12888-018-1679-x

33. Taylor RW, Marwood L, Oprea E, et al. Pharmacological Augmentation in Unipolar Depression: A Guide to the Guidelines. *Int J Neuropsychopharmacol*. 2021. doi:10.1093/ijnp/pyaa033

34. Strawbridge R, Carter B, Marwood L, et al. Augmentation therapies for treatment-resistant depression: Systematic review and meta-analysis. *Br J Psychiatry*. 2019. doi:10.1192/bjp.2018.233

35. EU-PEARL webpage. https://eu-pearl.eu/.

36. Saville BR, Berry SM. Efficiencies of platform clinical trials: A vision of the future. *Clin Trials*. 2016. doi:10.1177/1740774515626362

37. Woodcock J, LaVange LM. Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both. *N Engl J Med*. 2017. doi:10.1056/nejmra1510062

38. Angus DC, Alexander BM, Berry S, et al. Adaptive platform trials: definition, design, conduct and reporting considerations. *Nat Rev Drug Discov*. 2019. doi:10.1038/s41573-019-0034-3

39. J. JE, Linstone HA, Turoff M. The Delphi Method: Techniques and Applications. *Technometrics*. 1976. doi:10.2307/1268751

40. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010. doi:10.1371/journal.pmed.1000217

41. Jorm AF. Using the Delphi expert consensus method in mental health research. *Aust N Z J Psychiatry*. 2015. doi:10.1177/0004867415600891

42. Banno M, Tsujimoto Y, Kataoka Y. The majority of reporting guidelines are not developed with the Delphi method: a systematic review of reporting guidelines. *J Clin Epidemiol*. 2020. doi:10.1016/j.jclinepi.2020.04.010

43. Food and Drug Administration. *Major Depressive Disorder: Developing Drugs for Treatment, Guidance for Industry', DRAFT GUIDANCE. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Revision 1*.; 2018.

44. European Medicines Agency. *Guideline on Clinical Investigation of Medicinal Products in the Treatment of Depression. EMA/CHMP/185423/2010 Rev 2.*; 2013.

45. Jackson WC, Papakostas GI, Rafeyan R, Trivedi MH. Recognizing Inadequate Response in Patients With Major Depressive Disorder. *J Clin Psychiatry*. 2020. doi:10.4088/JCP.OT19037BR2

46. Hirschfeld RMA, Montgomery SA, Aguglia E, et al. Partial response and nonresponse to antidepressant therapy: Current approaches and treatment options. *J Clin Psychiatry*. 2002.

doi:10.4088/JCP.v63n0913

47. Kennedy SH, Lam RW, McIntyre RS, 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. *Can J Psychiatry*. 2016. doi:10.1177/0706743716659417

48. Thase ME, Trivedi MH, Nelson JC, et al. Examining the efficacy of adjunctive aripiprazole in major depressive disorder: A pooled analysis of 2 studies. *Prim Care Companion J Clin Psychiatry*. 2008. doi:10.4088/PCC.v10n0603

49. Daly EJ, Singh JB, Fedgchin M, et al. Efficacy and safety of intranasal esketamine adjunctive to oral antidepressant therapy in treatment-resistant depression: A randomized clinical trial. *JAMA Psychiatry*. 2018. doi:10.1001/jamapsychiatry.2017.3739

50. Buckman JEJ, Underwood A, Clarke K, et al. Risk factors for relapse and recurrence of depression in adults and how they operate: A four-phase systematic review and meta-synthesis. *Clin Psychol Rev*. 2018. doi:10.1016/j.cpr.2018.07.005

51. Fava GA, Offidani E. The mechanisms of tolerance in antidepressant action. *Prog Neuro-Psychopharmacology Biol Psychiatry*. 2011. doi:10.1016/j.pnpbp.2010.07.026

52. Voineskos D, Daskalakis ZJ, Blumberger DM. Management of treatment-resistant depression: Challenges and strategies. *Neuropsychiatr Dis Treat*. 2020. doi:10.2147/NDT.S198774

53. Luo Y, Kataoka Y, Ostinelli EG, Cipriani A, Furukawa TA. National Prescription Patterns of Antidepressants in the Treatment of Adults With Major Depression in the US Between 1996 and 2015: A Population Representative Survey Based Analysis. *Front Psychiatry*. 2020. doi:10.3389/fpsyt.2020.00035

54. Frazer A, Blier P. A Neuroscience-based Nomenclature (NbN) for psychotropic agents. *Int J Neuropsychopharmacol*. 2016. doi:10.1093/ijnppy/pyw066

55. Cantù F, Ciappolino V, Enrico P, Moltrasio C, Delvecchio G, Brambilla P. Augmentation with Atypical Antipsychotics for Treatment-Resistant Depression. *J Affect Disord*. 2021. doi:10.1016/j.jad.2020.11.006

56. Carter B, Strawbridge R, Husain MI, et al. Relative effectiveness of augmentation treatments for treatment-resistant depression: a systematic review and network meta-analysis. *Int Rev Psychiatry*. 2020. doi:10.1080/09540261.2020.1765748

57. 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 J Biol Psychiatry*. 2015. doi:10.3109/15622975.2014.1001786

58. Cleare A, Pariante CM, Young AH, et al. Evidence-based guidelines for treating depressive disorders with antidepressants: A revision of the 2008 British Association for Psychopharmacology guidelines. *J Psychopharmacol*. 2015. doi:10.1177/0269881115581093

59. 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 Scand*. 2020. doi:10.1111/acps.13145

60. 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. doi:10.1038/tp.2016.104

61. 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. doi:10.1176/appi.ajp.2015.15030331

62. 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. doi:10.3109/15622970701432528

63. Papakostas GI, Fava M, Thase ME. Treatment of SSRI-Resistant Depression: A Meta-Analysis Comparing Within- Versus Across-Class Switches. *Biol Psychiatry*. 2008. doi:10.1016/j.biopsych.2007.08.010

64. Bauer M, El-Khalili N, Datto C, Szamosi J, Eriksson H. A pooled analysis of two randomised, placebo-controlled studies of extended release quetiapine fumarate adjunctive to antidepressant therapy in patients with major depressive disorder. *J Affect Disord*. 2010. doi:10.1016/j.jad.2010.08.032

65. Brunner E, Tohen M, Osuntokun O, Landry J, Thase ME. Efficacy and Safety of Olanzapine/Fluoxetine Combination vs Fluoxetine Monotherapy Following Successful Combination Therapy of Treatment-Resistant Major Depressive Disorder. *Neuropsychopharmacology*. 2014. doi:10.1038/npp.2014.101

66. Marcus RN, McQuade RD, Carson WH, et al. The efficacy and safety of aripiprazole as adjunctive therapy in major depressive disorder: A second multicenter, randomized, double-blind, placebo-controlled study. *J Clin Psychopharmacol*. 2008. doi:10.1097/JCP.0b013e31816774f9

67. Thase ME, Youakim JM, Skuban A, et al. Efficacy and safety of adjunctive brexpiprazole 2 mg in major depressive disorder: A phase 3, randomized, placebo-controlled study in patients with inadequate response to antidepressants. *J Clin Psychiatry*. 2015. doi:10.4088/JCP.14m09688

68. Mahmoud RA, Pandina GJ, Turkoz I, et al. Risperidone for treatment-refractory major depressive disorder: A randomized trial. *Ann Intern Med*. 2007. doi:10.7326/0003-4819-147-9-200711060-00003

69. Papakostas GI, Fava M, Baer L, et al. Ziprasidone augmentation of escitalopram for major depressive disorder: Efficacy results from a randomized, Double-Blind, Placebo-Controlled Study. *Am J Psychiatry*. 2015. doi:10.1176/appi.ajp.2015.14101251

70. Nelson JC, Papakostas GI. Atypical antipsychotic augmentation in major depressive disorder: A meta-analysis of placebo-controlled randomized trials. *Am J Psychiatry*. 2009. doi:10.1176/appi.ajp.2009.09030312

71. Bauer M, Adli M, Ricken R, Severus E, Pilhatsch M. Role of lithium augmentation in the management of major depressive disorder. *CNS Drugs*. 2014. doi:10.1007/s40263-014-0152-8

72. Van Bronswijk S, Moopen N, Beijers L, Ruhe HG, Peeters F. Effectiveness of psychotherapy

for treatment-resistant depression: A meta-analysis and meta-regression. *Psychol Med*. 2019. doi:10.1017/S003329171800199X

73. Gloster AT, Rinner MTB, Ioannou M, et al. Treating treatment non-responders: A meta-analysis of randomized controlled psychotherapy trials. *Clin Psychol Rev*. 2020. doi:10.1016/j.cpr.2019.101810

74. Debonnel G, Saint-André É, Hébert C, De Montigny C, Lavoie N, Blier P. Differential physiological effects of a low dose and high doses of venlafaxine in major depression. *Int J Neuropsychopharmacol*. 2007. doi:10.1017/S1461145705006413

75. Furukawa TA, Cipriani A, Cowen PJ, Leucht S, Egger M, Salanti G. Optimal dose of selective serotonin reuptake inhibitors, venlafaxine, and mirtazapine in major depression: a systematic review and dose-response meta-analysis. *The Lancet Psychiatry*. 2019. doi:10.1016/S2215-0366(19)30217-2

76. 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. doi:10.4088/JCP.07m03780

77. Kelley ME, Dunlop BW, Nemeroff CB, et al. Response rate profiles for major depressive disorder: Characterizing early response and longitudinal nonresponse. *Depress Anxiety*. 2018. doi:10.1002/da.22832

78. Rush AJ, Kraemer HC, Sackeim HA, et al. Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology*. 2006. doi:10.1038/sj.npp.1301131

79. Rush AJ, Wisniewski SR, Warden D, et al. Selecting among second-step antidepressant medication monotherapies: Predictive value of clinical, demographic, or first-step treatment features. *Arch Gen Psychiatry*. 2008. doi:10.1001/archpsyc.65.8.870

80. 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 N Z J Psychiatry*. 2019. doi:10.1177/0004867419835933

81. Trivedi MH, Thase ME, Fava M, et al. Adjunctive aripiprazole in major depressive disorder: Analysis of efficacy and safety in patients with anxious and atypical features. *J Clin Psychiatry*. 2008. doi:10.4088/JCP.v69n1211

82. Bandelow B, Bauer M, Vieta E, et al. Extended release quetiapine fumarate as adjunct to antidepressant therapy in patients with major depressive disorder: Pooled analyses of data in patients with anxious depression versus low levels of anxiety at baseline. *World J Biol Psychiatry*. 2014. doi:10.3109/15622975.2013.842654

83. Thase ME, Weiller E, Zhang P, Weiss C, McIntyre RS. Adjunctive brexpiprazole in patients with major depressive disorder and anxiety symptoms: Post hoc analyses of three placebo-controlled studies. *Neuropsychiatr Dis Treat*. 2019. doi:10.2147/NDT.S185815

84. Papakostas GI, Petersen TJ, Farabaugh AH, et al. Psychiatric Comorbidity as a Predictor of Clinical Response to Nortriptyline in Treatment-Resistant Major Depressive Disorder. *J Clin Psychiatry*. 2003. doi:10.4088/JCP.v64n1112

85. Brenner P, Brandt L, Li G, DiBernardo A, Bodén R, Reutfors J. Substance use disorders and risk for treatment resistant depression: a population-based, nested case-control study. *Addiction*. 2020. doi:10.1111/add.14866

86. Nunes E V., Levin FR. Treatment of Depression in Patients with Alcohol or Other Drug Dependence: A Meta-analysis. *J Am Med Assoc*. 2004. doi:10.1001/jama.291.15.1887

87. Newton-Howes G, Tyrer P, Johnson T. Personality disorder and the outcome of depression: Meta-analysis of published studies. *Br J Psychiatry*. 2006. doi:10.1192/bjp.188.1.13

88. 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. doi:10.1177/0269881113499205

89. 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. doi:10.1016/j.euroneuro.2015.06.007

90. Pan LA, Martin P, Zimmer T, et al. Neurometabolic disorders: Potentially treatable abnormalities in patients with treatment-refractory depression and suicidal behavior. *Am J Psychiatry*. 2017. doi:10.1176/appi.ajp.2016.15111500

91. Regier DA, Narrow WE, Clarke DE, et al. DSM-5 field trials in the United States and Canada, part II: Test-retest reliability of selected categorical diagnoses. *Am J Psychiatry*. 2013. doi:10.1176/appi.ajp.2012.12070999

92. Taylor, D.M., Barnes, T.R. and Young A. *The Maudsley Prescribing Guidelines in Psychiatry*. John Wiley & Sons.; 2018.

93. Cheruvu VK, Chiyaka ET. Prevalence of depressive symptoms among older adults who reported medical cost as a barrier to seeking health care: findings from a nationally representative sample. *BMC Geriatr*. 2019. doi:10.1186/s12877-019-1203-2

94. Yalin N, Young AH. The age of onset of unipolar depression. In: *Age of Onset of Mental Disorders: Etiopathogenetic and Treatment Implications*. ; 2018. doi:10.1007/978-3-319-72619-9_6

95. First, M. B., Spitzer, R.L, Gibbon M., and Williams JB. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders*.; 2002.

96. First MB, Williams JBW, Karg RS, Spitzer RL. Structured clinical interview for DSM-5 research version. *Am Psychiatr Assoc Washingt DC*. 2015.

97. Sheehan D V., Lecrubier Y, Sheehan KH, 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.

98. Chandler GM, Iosifescu D V., Pollack MH, Targum SD, Fava M. Validation of the massachusetts general hospital Antidepressant Treatment History Questionnaire (ATRQ). *CNS Neurosci Ther*. 2010. doi:10.1111/j.1755-5949.2009.00102.x

99. Sackeim HA. The definition and meaning of treatment-resistant depression. *J Clin Psychiatry*. 2001.

100. Desseilles M, Witte J, Chang TE, et al. Assessing the adequacy of past antidepressant trials: A clinician's guide to the antidepressant treatment response questionnaire. *J Clin Psychiatry*. 2011. doi:10.4088/JCP.11ac07225

101. Sackeim HA, Aaronson ST, Bunker MT, et al. The assessment of resistance to antidepressant treatment: Rationale for the Antidepressant Treatment History Form: Short Form (ATHF-SF). *J Psychiatr Res*. 2019. doi:10.1016/j.jpsychires.2019.03.021

102. Ruhé HG, Van Rooijen G, Spijker J, Peeters FPML, Schene AH. Staging methods for treatment resistant depression. A systematic review. *J Affect Disord*. 2012. doi:10.1016/j.jad.2011.02.020

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

104. Fekadu A, Wooderson S, Donaldson C, et al. A multidimensional tool to quantify treatment resistance in depression: The Maudsley staging method. *J Clin Psychiatry*. 2009. doi:10.4088/JCP.08m04309

105. Van Belkum SM, Geugies H, Lysen TS, et al. Validity of the maudsley staging method in predicting treatment-resistant depression outcome using the netherlands study of depression and anxiety. *J Clin Psychiatry*. 2018. doi:10.4088/JCP.17m11475

106. HAMILTON M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960. doi:10.1136/jnnp.23.1.56

107. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979. doi:10.1192/bjp.134.4.382

108. Rush AJ, Trivedi MH, Ibrahim HM, 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. doi:10.1016/S0006-3223(02)01866-8

109. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An Inventory for Measuring Depression. *Arch Gen Psychiatry*. 1961. doi:10.1001/archpsyc.1961.01710120031004

110. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med*. 2001. doi:10.1046/j.1525-1497.2001.016009606.x

111. Guy W. Rating Clinician-rated | Clinical Global Impression (CGI). *ECDEU Assess Man Psychopharmacol*. 1976.

112. DVRush AS, Pincus H, First M. Sheehan disability scale. *Handb Psychiatr Meas*. 2000.

113. Deshpande P, Sudeepthi Bl, Rajan S, Abdul Nazir C. Patient-reported outcomes: A new era in clinical research. *Perspect Clin Res*. 2011. doi:10.4103/2229-3485.86879

114. Frank E, Prien RF, Jarrett RB, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recovery, relapse, and recurrence. *Arch Gen Psychiatry*. 1991. doi:10.1001/archpsyc.1991.01810330075011

115. Hawley CJ, Gale TM, Sivakumaran T. Defining remission by cut off score on the MADRS: Selecting the optimal value. *J Affect Disord*. 2002. doi:10.1016/S0165-0327(01)00451-7

116. Zimmerman M, Posternak MA, Chelminski I. Defining remission on the montgomery-asberg

depression rating scale. *J Clin Psychiatry*. 2004. doi:10.4088/JCP.v65n0204

117. Uher R, Perlis RH, Placentino A, et al. Self-report and clinician-rated measures of depression severity: Can one replace the other? *Depress Anxiety*. 2012. doi:10.1002/da.21993

118. Dunlop BW, Li T, Kornstein SG, et al. Concordance between clinician and patient ratings as predictors of response, remission, and recurrence in major depressive disorder. *J Psychiatr Res*. 2011. doi:10.1016/j.jpsychires.2010.04.032

119. Fried EI. The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *J Affect Disord*. 2017. doi:10.1016/j.jad.2016.10.019

120. HAMILTON M. THE ASSESSMENT OF ANXIETY STATES BY RATING. *Br J Med Psychol*. 1959. doi:10.1111/j.2044-8341.1959.tb00467.x

121. Carhart-Harris RL, Bolstridge M, Rucker J, et al. Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. *The Lancet Psychiatry*. 2016. doi:10.1016/S2215-0366(16)30065-7

122. Felger JC, Li Z, Haroon E, et al. Inflammation is associated with decreased functional connectivity within corticostriatal reward circuitry in depression. *Mol Psychiatry*. 2016. doi:10.1038/mp.2015.168

123. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH Guidelines for themanagement of arterial hypertension. *Eur Heart J*. 2018. doi:10.1093/eurheartj/ehy339

124. Cattaneo A, Ferrari C, Turner L, et al. Whole-blood expression of inflammasome- and glucocorticoid-related mRNAs correctly separates treatment-resistant depressed patients from drug-free and responsive patients in the BIODEP study. *Transl Psychiatry*. 2020. doi:10.1038/s41398-020-00874-7

125. Wittenberg GM, Greene J, Vértes PE, Drevets WC, Bullmore ET. Major Depressive Disorder Is Associated With Differential Expression of Innate Immune and Neutrophil-Related Gene Networks in Peripheral Blood: A Quantitative Review of Whole-Genome Transcriptional Data From Case-Control Studies. *Biol Psychiatry*. 2020. doi:10.1016/j.biopsych.2020.05.006

126. Lynall ME, Turner L, Bhatti J, et al. Peripheral Blood Cell–Stratified Subgroups of Inflamed Depression. *Biol Psychiatry*. 2020. doi:10.1016/j.biopsych.2019.11.017

127. Uher R, Tansey KE, Dew T, et al. An inflammatory biomarker as a differential predictor of outcome of depression treatment with escitalopram and nortriptyline. *Am J Psychiatry*. 2014. doi:10.1176/appi.ajp.2014.14010094

128. Chamberlain SR, Cavanagh J, De Boer P, et al. Treatment-resistant depression and peripheral C-reactive protein. *Br J Psychiatry*. 2019. doi:10.1192/bjp.2018.66

129. Nouraei H, Firouzabadi N, Mandegary A, et al. Glucocorticoid receptor genetic variants and response to fluoxetine in major depressive disorder. *J Neuropsychiatry Clin Neurosci*. 2018. doi:10.1176/appi.neuropsych.16120322

130. O'Connell CP, Goldstein-Piekarski AN, Nemeroff CB, et al. Antidepressant outcomes predicted by Genetic variation in corticotropin-releasing hormone binding protein. *Am J Psychiatry*. 2018. doi:10.1176/appi.ajp.2017.17020172

131. Uhr M, Tontsch A, Namendorf C, et al. Polymorphisms in the Drug Transporter Gene ABCB1 Predict Antidepressant Treatment Response in Depression. *Neuron*. 2008. doi:10.1016/j.neuron.2007.11.017

132. Porcelli S, Fabbri C, Serretti A. Meta-analysis of serotonin transporter gene promoter polymorphism (5-HTTLPR) association with antidepressant efficacy. *Eur Neuropsychopharmacol*. 2012. doi:10.1016/j.euroneuro.2011.10.003

133. Murphy GM, Kremer C, Rodrigues HE, Schatzberg AF. Pharmacogenetics of antidepressant medication intolerance. *Am J Psychiatry*. 2003. doi:10.1176/appi.ajp.160.10.1830

134. Anttila S, Huuhka K, Huuhka M, et al. Interaction between 5-HT1A and BDNF genotypes increases the risk of treatment-resistant depression. *J Neural Transm*. 2007. doi:10.1007/s00702-007-0705-9

135. Akil H, Gordon J, Hen R, et al. Treatment resistant depression: A multi-scale, systems biology approach. *Neurosci Biobehav Rev*. 2018. doi:10.1016/j.neubiorev.2017.08.019

136. Krystal AD, Pizzagalli DA, Mathew SJ, et al. The first implementation of the NIMH FAST-FAIL approach to psychiatric drug development. *Nat Rev Drug Discov*. 2018. doi:10.1038/nrd.2018.222

137. Chevance A, Ravaud P, Tomlinson A, et al. Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey. *The Lancet Psychiatry*. 2020. doi:10.1016/S2215-0366(20)30191-7

138. Mäntylä FL. Major Depressive Disorder - A Patient Perspective to Recovery. https://www.inspirethemind.org/blog/major-depressive-disorder-a-patient-perspective-to-recovery.

139. 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. doi:10.1007/s40263-017-0417-0

140. ten Doesschate MC, Bockting CLH, Schene AH. Adherence to continuation and maintenance antidepressant use in recurrent depression. *J Affect Disord*. 2009. doi:10.1016/j.jad.2008.07.011

141. 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. doi:10.1016/j.jad.2015.12.029

142. Holma IAK, Holma KM, Melartin TK, Isometsä ET. Treatment attitudes and adherence of psychiatric patients with major depressive disorder: A five-year prospective study. *J Affect Disord*. 2010. doi:10.1016/j.jad.2010.04.022

143. Laursen DRT, Paludan-Müller AS, Hróbjartsson A. Randomized clinical trials with run-in periodsfrequency, characteristics and reporting. *Clin Epidemiol*. 2019. doi:10.2147/CLEP.S188752

Annexes

Annex I: Tables

Table 1: Overview of current definitions of treatment-resistance in MDD.

	FDA (2018)	*EMA (2013)*	*Salloum and Papakostas (2019)*	*Gaynes et al. (2020)*	*McAllister-Williams et al. (2020)*	*Identified issues - problems*
Previous AD courses	No response to *more than one* prior antidepressant.	At least *two failed trials.*	A *continuous spectrum* starting with *one failed trial.*	A minimum of *two prior treatment failures.*	At least *two treatment trials* (*one* in some cases).	Different criteria for TRD-PRD in regulatory statements and clinical studies.
AD dose	*Adequate dose* (not specified).	*Adequate dose* (not specified).	*Adequate dose* (or at minimal plasma levels). There should be a differentiation between minimally effective treatment and optimal treatment. It should be clearly defined from the onset.	Confirmation of *prior adequate dose* (not specified).	Therapeutic dosage *adequate for target engagement.*	What is an adequate dosage (maximum licensed dosage? maximum dose tolerated? minimally effective dose? lowest approved dose? other?). ADs have wide therapeutic index.
Treatment duration	*Adequate duration* (not specified).	*Sufficient length of time* (not specified) for TRD definition; 4-6 weeks for clinical trials in TRD.	*Sufficient duration* (not specified). It should be clearly defined from the onset.	Confirmation of *prior adequate* duration (*minimum 4 weeks*).	Doses should be *maintained* (not specified for how long).	For how long should a specific treatment be maintained?

	FDA (2018)	EMA (2013)	Salloum and Papakostas (2019)	Gaynes et al. (2020)	McAllister-Williams et al. (2020)	Identified issues - problems
Type of previous treatments	Not specified.	Antidepressants of *same or different class.*	Not specified. Augmentation therapies, psychotherapy and other evidence-based brain stimulation therapies should be reflected when calculating treatment strength. Treatment with ECT or ketamine should be ascribed higher scores.	Any intervention tested and identified as a treatment for TRD: pharmacologic treatments, non-pharmacologic devices or procedures, psychotherapy, complementary and alternative medicine.	Past treatment trials, whether psychological, pharmacological or neurostimulatory. The types of medication previously used may inform subsequent treatment choices.	Which drugs count as efficacious besides SSRI/SNRI? Other interventions? Psychotherapy?
Recommended tools	Primary endpoint: HAMD17, MADRS. Secondary endpoint: CGI, SDS.	Primary endpoint: HAMD (preferably the 17-item), and the MADRS. Secondary endpoint: changes in global assessment (e.g., CGI) or in social functioning.	Validated scales (not specified).	Patient-reported instruments to be preferred.	Both symptoms rating scales and measures of psychosocial functions.	Different criteria for TRD-PRD patients? Different scales more sensitive to change?

Table 2: Main consensus recommendations on TRD/PRD regulatory clinical trials.

Consensus key points		
Recommendations which can be implemented immediately within current practice		***Level of consensus***
TRD and PRD definitions		
1	A definition of TRD for clinical trials conducted for regulatory purposes is necessary.	*Strong*
2	A definition of PRD – as a distinct group from TRD – for clinical trials conducted for regulatory purposes is recommended.	*Moderate*
Previous antidepressant treatments		
3	Failing at least two prior antidepressant treatments – in the current episode – is necessary to define TRD.	*Strong*
4	A partial response to one single antidepressant treatment – in the current episode – is sufficient to define PRD.	*Moderate*
5	There is no maximum number of previous antidepressant treatments for patients recommended to be included in TRD (and PRD) regulatory trials, but all documentable life-time treatment response should be recorded to define the TRD (and PRD) stage of a patient *(see also below)*.	*Moderate*
6	TRD (and PRD) definition (and differentiation) should be based on the current depressive episode only, and in the past two years only; if the current episode has lasted more than two years, treatments prior to the last two years should not be considered.	*Weak*
7	To define TRD (and PRD), response to previous antidepressant treatments (within the current episode and in the past two years) can be ascertained retrospectively using structured interviews and clinical documentation.	*Moderate*
Type, dose, and duration of an 'adequate' antidepressant treatment trial		
8	To define TRD, the two different treatment failures must involve two established medications for MDD, with different mechanisms of action.	*Moderate*
9	Regulatory clinical trials for TRD (and PRD) may include patients who failed to respond (or partially responded) to augmentation/combination treatment strategies, but these need to be primarily based on medical records.	*Strong*
10	Regulatory clinical trials for TRD (and PRD) may include patients who failed to respond (or partially responded) to brain stimulation treatments, such as transcranial magnetic stimulation (TMS) and electroconvulsive therapy (ECT).	*Moderate*
11	Regulatory clinical trials for TRD (and PRD) should not include patients who failed to respond (or partially responded) to vagus nerve stimulation (VNS) and deep brain stimulation (DBS).	*Moderate*
12	Regulatory clinical trials for TRD (and PRD) may include patients who failed to respond (or partially responded) to structured psychotherapy.	*Strong*
13	The minimum effective dose of a medication indicated in MDD is enough to define *a treatment failure for the purpose of establishing* TRD (and PRD).	*Moderate*
14	For conventional medications for MDD, a treatment given for at least four weeks in duration is sufficient to define *a treatment failure for the purpose of establishing* TRD (and PRD).	*Moderate*

Consensus key points		
Recommendations which can be implemented immediately within current practice		***Level of consensus***
15	Patients' discontinuation of treatment before the completion of the fourth week, should not be considered as *a treatment failure for the purpose of establishing* TRD (and PRD).	*Strong*
Clinical presentation of TRD and PRD patients		
16	No specific types of symptoms of MDD should either be prerequisite, or excluded, from the definition of TRD (and PRD), but symptoms and specifiers should always be recorded.	*Strong*
17	Patients with bipolar depression should be excluded from TRD (and PRD) studies, as this is a separate condition from unipolar depression (MDD).	*Strong*
18	Patients with comorbid personality disorders or other mental disorders should be excluded from TRD (and PRD) studies only when their onset is properly documented as independent and antecedent to the MDD diagnosis.	*Moderate*
19	Patients with comorbid substance use disorder that is active and severe should always be excluded from TRD (and PRD) studies, independently from the onset; *in contrast*, patients with comorbid substance use disorder that is active and mild/moderate should be excluded from TRD (and PRD) studies only when the onset is properly documented as independent and antecedent to the MDD diagnosis.	*Moderate*
Diagnostic tools and measures of outcome		
20	Maudsley Staging Model is the suggested instrument to define the degree of treatment resistance historically.	*Moderate*
21	Clinician administered MADRS10 is the suggested outcome instrument to assess treatment response (and remission) and, together with patient-reported QIDS-SR, can be used to assess TRD and PRD status.	*Moderate*
22	Criteria for remission, response, and partial response should not be relaxed in regulatory clinical trials for TRD (and PRD); shorter versions of the traditional scales, such as the HAMD6 and the MADRS6, should not be currently preferred to traditional scales, although may become more relevant in the future with fast-acting interventions.	*Moderate*
Recommendations which can be implemented in future research		
23	Future research should be more patient-centred, recognizing, and targeting different clinical phenotypes of TRD and PRD underpinned by a specific biological mechanism.	*Strong*
24	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 patient.	*Strong*
25	Currently, no biomarker has been validated in clinical practice or in clinical trials to identify TRD (and PRD) patients, or to further stratify them; however, collection of biological samples for subsequent subgroup or stratified analyses is recommended.	*Moderate*
26	Patients' preferences, perspectives, and reported outcomes should be included in future TRD (and PRD) diagnostic tools and outcome measures.	*Strong*
27	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*

Annex II: Initial questionnaire for experts

In order to gather experts' opinion on these gaps in knowledge, we have prepared the following questionnaire that we would kindly ask you to fill out. Please write as much or as little as you prefer, but please do try to answer all questions, even if just with a yes or a no. Please also remember that your answers will be ultimately used to improve the quality of future regulatory clinical trials in PRD-TRD, so keep this priority in mind.

1. Do you think that it is useful to have a definition of TRD for clinical trials conducted for regulatory purposes?
2. Do you think that it is useful to differentiate between TRD and PRD? If so, should it be based only on history (response to previous antidepressants) or also on current depressive symptoms while on antidepressants? How can patients' preferences and attitudes be included in this definition?
3. How many failed prior treatment trials do you think are necessary to define PRD-TRD? Should the number of failed trials include the current episode? If it is two or more, should at least one antidepressant trial be done prospectively as part of a regulatory clinical trial, before testing the novel compound? Should a maximum number of failed previous treatment for patients be included in regulatory trials, to select a group which is still amenable to some improvement?
4. Do you think that some types of symptoms should either be prerequisite, o should be excluded, from the definition of TRD? For example, should the presence of melancholic, atypical or anxious symptoms be only recorded, or should it be used to guide inclusion or exclusion? How about comorbidity with personality disorders or substance abuse?
5. Are there any genetic markers that you would suggest in the identification of TRD patients, or to further stratify them? Are these tests ready to be used currently (sufficient validity, sensitivity, and specificity)?
6. Which dosage should be considered when defining a treatment failure, the minimal approved dosage or the maximum approved dosage? Or something else? How do we assess compliance? Should medications level be conducted at least to confirm current treatment failure?
7. How long do you think a treatment trial must last before considering it a failure? At least 4 weeks? At least 8 weeks? More?
8. Which antidepressants history scale or staging model would you use to define PRD or TRD, among the ones discussed above?
9. Which depressive symptoms scales and score would you use to define PRD or TRD in someone who is currently on antidepressants? Should the core symptoms scales (like HDRS6 and the MADRS6) be used instead of the full scales?
10. Should the criteria for full/partial response and for remission (for example, 50% drop in score or HAMD<7) be relaxed in regulatory clinical trial for PRD-TRD, to allow for the fact that these patients are unlikely to show full improvement even with new medications?
11. Should the definition of TRD for regulatory purpose of clinical trials include failure to augmentation/combination of antidepressants? What about failure to psychotherapeutic

interventions or brain stimulation? Or should patients who have failed at these strategies be excluded from PRD-TRD regulatory trials, to minimize the variability of the study population and maximizing the chances of identify a therapeutic effect?

Annex III: Experts Meeting participants and schedule

EU-PEARL WP4

Consensus Meeting

22nd May 2020

Morning session: 09^{00}-12^{00} BST / 04^{00}-07^{00} EDT / 01^{00}-04^{00} PDT

Afternoon session: 15^{00}-18^{00} BST / 10^{00}-13^{00} EDT / 07^{00}-10^{00} PDT

Treatment-response in Major Depressive Disorder

Participants

International experts:			
Morning session	Ian Anderson	*Afternoon session*	Pierre Blier
	Bruno Aouizerate		Ulrich Hegerl
	Volker Arolt		Andrew D. Krystal
	Bernhard Baune		Andrew H. Miller
	Michael Bauer		Charles B. Nemeroff
	Anthony Cleare		David Nutt
	Philip Cowen		Claus Normann
	Ted Dinan		Stefano Pallanti
	Andrea Fagiolini		Luca Pani
	Nicol Ferrier		Roger McIntyre
	Marion Leboyer		Allan F. Schatzberg
	Hamish McAllister-Williams		Richard C. Shelton
	Andreas Meyer-Lindenberg		Lakshmi Yatham
	Brenda Penninx		Roland Zahn
	Allan Young		

EU-PEARL WP4 leaders:

Yanina Flossbach
Stefan Gold
Eduard Maron
Christian Otte
Carmine M. Pariante

Josep Antoni Ramos-Quiroga
Adam Savitz

EU-PEARL WP4 contributors:

Francesco Benedetti
Gara Arteaga Henriquez
Witte Hoogendijk
Heddie Martynowicz
Mark Schmidt
Luca Sforzini
Courtney Worrell

Schedule (BST)	**Session**	**Chair**	**Speakers**
9^{00}-9^{15} *15^{00}-15^{15}*	Introduction		Carmine M. Pariante
9^{15}-9^{25} *15^{15}-15^{25}*	1st topic - *Is it helpful to separate TRD vs PRD or is it all a continuous spectrum?*	Carmine M. Pariante	Bernhard Baune & Ted Dinan Charles B. Nemeroff & Richard C. Shelton
9^{25}-9^{45} *15^{25}-15^{45}*	1st topic discussion		Whole group
9^{45}-9^{55} *15^{45}-15^{55}*	2nd topic – *Can we really assess treatment failure historically, beyond the current episode?*	Christian Otte Stefan M. Gold	Nicol Ferrier & Brenda Penninx Roger McIntyre & Luca Pani
9^{55}-10^{15} *15^{55}-16^{15}*	2nd topic discussion		Whole group
10^{15}-10^{30} *16^{15}-16^{30}*	Future research - *Clinical practice vs. Regulatory needs*	Carmine M. Pariante	Ian Anderson Lakshmi Yatham
10^{30}-10^{45} *16^{30}-16^{45}*	Break		
10^{45}-10^{55} *16^{45}-16^{55}*	3rd topic – *Is 4 weeks of AD at the minimal therapeutic dose enough to demonstrate treatment resistance?*	Josep Antoni Ramos-Quiroga Adam Savitz	Philip Cowen & Michael Bauer Ulrich Hegerl & David Nutt
10^{55}-11^{15} *16^{55}-17^{15}*	3rd topic discussion		Whole group
11^{15}-11^{25} *17^{15}-17^{25}*	4th topic - *Is TRD comorbid with personality disorder or substance abuse a different*	Yanina Flossbach	Allan Young & Anthony Cleare

	psychiatric disorder?	Eduard Maron	Pierre Blier & Alan F. Schatzberg
11^{25}-11^{45} *17^{25}-17^{45}*	4th topic discussion		Whole group
11^{45}-12^{00} *17^{45}-18^{00}*	Future research – *Cluster of symptoms and biomarkers*	Carmine M. Pariante	Andrea Fagiolini Andrew H. Miller

EU-PEARL WP4

Stakeholder Meeting

9th October 2020

15^{00}-17^{00} BST (London time) / 10^{00}-12^{00} EDT (New York time)

Treatment-response in Major Depressive Disorder

Schedule (BST)	**Session**
15^{00}-15^{10}	Introduction
15^{10}-15^{35}	*EU-PEARL project and main objectives in MDD – defining TRD and PRD*
15^{35}-15^{50}	*Operational criteria for TRD and PRD definitions*
15^{50}-16^{15}	Discussion – *Whole group*
16^{15}-16^{30}	*Future clinical trials and platform trials in TRD and PRD*
16^{30}-16^{55}	Discussion – *Whole group*
16^{55}-17^{00}	Concluding remarks

Participants

Name	**Affiliation**
Georgios Aistlaitner	BfArM
Alun Bedding	Roche
Francesco Benedetti	HSR
Roberto Furlan	HSR
Stefan Gold	Charité
Gara Harteaga Henriquez	VHIR
Anne Kaminski	Janssen
Melisa Kose	KCL
Fanni-Laura Mäntylä	GAMIAN
Jadwiga Martynowicz	Janssen

Elisa Melloni	HSR
Christina Müller	TriNetX
Carmine Maria Pariante	KCL
Adam Savitz	Janssen
Mark Schmidt	Janssen
Luca Sforzini	KCL
Cécile Spiertz	Janssen
Andrew Thomson	EMA
Edwin van de Ketterij	EATRIS
Courtney Worrell	KCL

Annex IV: Disclaimer

This document reflects the majority views of the experts (See
). Authorship reflects having been part of the process and accepting the consensus statements without necessarily personally endorsing every single recommendation (see level of consensus of each recommendation). 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 deliverable reflects only the authors' view and the JU is not responsible for any use that may be made of the information it contains.

Annex V: Full list of contributors to D4.1 and acknowledgement

EU-PEARL members: Luca Sforzini[1], Courtney Worrell[1], Melisa Kose[1], Gara Arteaga-Henríquez[37,38,39], Francesco Benedetti[40,41], Stefan M. Gold[42,43,44], Witte J. G. Hoogendijk[45], Eduard Maron[46], Heddie Martynowicz[47], Christian Otte[42], Mark E. Schmidt[48], Edwin van de Ketterij[49], Katherine Woo[47], Yanina Flossbach[50], J. Antoni Ramos-Quiroga[37,39,51], Adam J. Savitz[52], and Carmine M. Pariante[1,9].

External contributors: Ian M. Anderson[2], Bruno Aouizerate[3], Volker Arolt[4,5], Michael Bauer[6], Bernhard T. Baune[4,7], Pierre Blier[8], Anthony Cleare[1,9], Philip J. Cowen[10], Timothy G. Dinan[11], Andrea Fagiolini[12], I. Nicol Ferrier[13,14], Ulrich Hegerl[15], Andrew D. Krystal[16], Marion Leboyer[17], R. Hamish McAllister-Williams[13,14], Roger S. McIntyre[18], Andreas Meyer-Lindenberg[19], Andrew H. Miller[20], Charles B. Nemeroff[21], Claus Normann[22], David Nutt[23], Stefano Pallanti[24], Luca Pani[25], Brenda W. J. H. Penninx[26], Alan F. Schatzberg[27], Richard C. Shelton[28], Lakshmi N. Yatham[29], Allan H. Young[1,9], Roland Zahn[1,9], Georgios Aislaitner[30], Florence Butlen-Ducuing[31], Christine Fletcher[32], Marion Haberkamp[30], Thomas Laughren[33], Fanni-Laura Mäntylä[34], Koen Schruers[35], Andrew Thomson[36].

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; Department of Psychiatry, University of Ottawa, Ottawa, ON, Canada; Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada. [9]National Institute for Health Research Mental Health Biomedical Research Centre, South London and Maudsley NHS Foundation Trust and King's College London, London, UK. [10]Medical Sciences Division, Department of Psychiatry, University of Oxford, Oxford, UK. [11]APC Microbiome Ireland, Cork, Ireland; Department of Psychiatry and Neurobehavioral Sciences, University College Cork, Cork, Ireland. [12]Department of Molecular Medicine, Division of Psychiatry, University of Siena, Siena, Italy. [13] Translational and Clinical

Research Institute, Newcastle University, Newcastle upon Tyne, UK. [14]Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust, Newcastle, UK. [15]Depression Research Center of the German Depression Foundation and Department of Psychiatry, Psychosomatics and Psychotherapy, Goethe University, Frankfurt, Germany. [16]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA; Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, USA. [17]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. [18]Department of Psychiatry, University of Toronto, Toronto, ON, Canada; Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON, Canada; Mood Disorders Psychopharmacology Unit, University Health Network, Toronto, ON, Canada; Brain and Cognition Discovery Foundation, Toronto, ON, Canada. [19]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, 68159, Mannheim, Germany. [20]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, 30322, USA. [21]Department of Psychiatry, University of Texas at Austin, Dell Medical School, Austin, TX, USA. [22]Department for Psychiatry and Psychotherapy, Medical Center – University of Freiburg, Faculty of Medicine University of Freiburg, Hauptstrasse 5, 79104, Freiburg, Germany. [23]Centre for Neuropsychopharmacology, Division of Psychiatry, Imperial College, London, UK. [24]Istituto di Neuroscience, University of Florence, Italy; Albert Einstein College of Medicine, New York, USA. [25]Department of Psychiatry and Behavioral Sciences, Psychiatry University of Miami, Miami, FL, USA; Department of Biomedical, Metabolic & Neural Sciences, University of Modena, Modena, Italy; VeraSci, Durham, NC, USA. [26]Department of Psychiatry, Amsterdam UMC, Vrije Universiteit and GGZinGeest, Amsterdam, the Netherlands. [27]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [28]Department of Psychiatry, University of Alabama at Birmingham, Birmingham, AL, USA. [29]Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada.
[30]Federal Institute for Drugs and Medical Devices (Bundesinstitut für Arzneimittel und Medizinprodukte, BfArM), Bonn, Germany. [31]Office of Therapies for Neurological and Psychiatric disorders, Human Medicines Division, European Medicines Agency, Amsterdam, the Netherlands. [32]Biostatistics, GlaxoSmithKline. [33]Laughren Psychopharm Consulting, LLC. [34]GAMIAN-Europe (Global Alliance of Mental Illness Advocacy Networks-Europe), Brussels, Belgium. [35]Department of Psychiatry and Psychology, School for Mental Health and Neuroscience, EURON, Maastricht University Medical Centre, Maastricht, the Netherlands; Faculty of Psychology, Center for Experimental and Learning Psychology, University of Leuven, Leuven, Belgium. [36]Data, Analytics and Methodology Taskforce, European Medicines Agency, Amsterdam, the Netherlands. [37]Department of Psychiatry, Hospital Universitari Vall d'Hebron (HUVH), Barcelona, Catalonia, Spain. [38]Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain. [39]Department of Psychiatry and Forensic Medicine, Universitat

Autònoma de Barcelona, Barcelona, Catalonia Spain. [40]Vita-Salute San Raffaele University, Milan, Italy. [41]Division of Neuroscience, Psychiatry and Clinical Psychobiology, IRCCS San Raffaele Scientific Institute, Milan, Italy. [42]Charité – Universitätsmedizin Berlin, Department of Psychiatry and Psychotherapy, Campus Benjamin Franklin, Berlin, Germany. [43]Charité – Universitätsmedizin Berlin, Department of Psychosomatic Medicine, Campus Benjamin Franklin, Berlin, Germany. [44]University Medical Center Hamburg-Eppendorf, Institute of Neuroimmunology and Multiple Sclerosis (INIMS), Center for Molecular Neurobiology, Hamburg, Germany. [45]Department of Psychiatry, Erasmus University Medical Centre, Rotterdam, the Netherlands. [46]Department of Psychiatry, University of Tartu, Tartu, Estonia; Faculty of Medicine, Department of Medicine, Centre for Neuropsychopharmacology, Division of Brain Sciences, Imperial College London, London, UK; Documental Ltd, Tallin, Estonia. [47]Quantitative Sciences, Janssen Research & Development, LLC, USA. [48]Experimental Medicine, Janssen Research & Development, Janssen Pharmaceutica NV, Beerse, Belgium. [49]European Infrastructure for Translational Medicine (EATRIS), Amsterdam, Netherlands. [50]Neuroscience, Global Drug Development, Novartis Pharma AG, Basel, Switzerland. [51]Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Catalonia Spain. Group of Psychiatry, Mental Health and Addictions, Vall d'Hebron Research Institute (VHIR), Barcelona, Catalonia, Spain; Biomedical Network Research Centre on Mental Health (CIBERSAM), Barcelona, Catalonia, Spain. [52]Department of Neuroscience, Janssen Research & Development, LLC, Titusville, NJ, USA.

LS and CMP are principal authors with full access to all of the data in this consensus report, and are responsible for writing all the different versions of the manuscript. CW, MK and all the authors involved within the EU-PEARL project (GA-H, FB, SMG, WJGH, EM, HM, CO, MES, EvdK, KW, YF, JAR-Q, AJS) contributed to the conception and design of the work, data acquisition, and initial drafting of the manuscript. All other authors contributed substantially to the writing of the manuscript and the consensus recommendations. All authors contributed to the critical revision of the manuscript for intellectual content and all are responsible for the content of this consensus paper.

Annex VI: Declaration of interests

LS reports grant from European Commission (EU-PEARL, IMI2 Grant Agreement no. 853966), during the conduct of the study.

CW reports grant from European Commission (EU-PEARL, IMI2 Grant Agreement no. 853966), during the conduct of the study.

MK reports grant from European Commission (EU-PEARL, IMI2 Grant Agreement no. 853966), during the conduct of the study.

IMA has nothing to disclose.

BA reports personal fees from Lundbeck, personal fees and non-financial support from Janssen-Cilag, personal fees from Sanofi, personal fees from Eli Lilly, outside the submitted work; and has served on the advisory board of Janssen-Cilag.

VA reports personal fees from Astra-Zeneca, personal fees from Gilead Sciences, personal fees from Janssen, personal fees from Lundbeck, personal fees from Neuraxpharm, personal fees from Otsuka, personal fees from Sanofi/Genzyme, personal fees from Servier, grants from European Union-Horizon 2020, grants from Medical Faculty Muenster- Interdisciplinary Center for Clinical Research, outside the submitted work; and Has served on Advisory Boards for Allergan, Astra-Zeneca, Janssen, Lundbeck, Neuraxpharm, Otsuka, Sanofi/Genzyme, Servier and Trommsdorff.

MB reports grants from Deutsche Forschungsgemeinschaft (DFG), grants from European Commission, grants from Bundesministerium für Bildung und Forschung (BMBF), personal fees from Aristo, personal fees from GH Research, personal fees from Hexal, personal fees from Janssen, personal fees from Janssen-Cilag, personal fees from Neuraxpharm, personal fees from Novartis, personal fees from Sandoz, personal fees from Shire International, personal fees from Sunovion, personal fees from Takeda, outside the submitted work.
BTB reports personal fees from AstraZeneca, personal fees from Lundbeck, personal fees from Pfizer, personal fees from Takeda, personal fees from Servier, personal fees from Bristol Myers Squibb, personal fees from Otsuka, personal fees from LivaNova, personal fees from Janssen-Cilag, outside the submitted work.

PB reports personal fees from Allergan, personal fees from Eisai, personal fees from Janssen, personal fees from Lundbeck, personal fees from Otsuka, personal fees from Pfizer, personal fees from Pierre Fabre Médicaments, personal fees from Takeda, outside the submitted work; and Expert testimony was provided on behalf of Allergan, Bristol Myers Squibb and Otsuka.
AJC reports personal fees from Lundbeck, personal fees from Janssen, personal fees from Livanova,

personal fees from Allergan, personal fees from NICE, grants from Medical Research Council (UK), grants from Wellcome Trust (UK), grants from National Institute for Health Research (UK), grants from Protexin Probiotics International Ltd., outside the submitted work.

PJC has a patent for Ebselen in treatment-resistant depression pending.

TGD has nothing to disclose.

AF reports grants and personal fees from Angelini, grants and personal fees from Apsen, grants and personal fees from Boheringer Ingelheim, grants and personal fees from Daiichi Sankyo, grants and personal fees from Doc Generici, grants and personal fees from Glaxo Smith Kline, grants and personal fees from Italfarmaco, grants and personal fees from Lundbeck , grants and personal fees from Janssen, grants and personal fees from Mylan, grants and personal fees from Neuraxpharm, grants and personal fees from Otsuka, grants and personal fees from Pfizer, grants and personal fees from Recordati, grants and personal fees from Sanofi Aventis, grants and personal fees from Sunovion, grants and personal fees from Vifor, outside the submitted work.

INF has nothing to disclose.

UH reports personal fees from Janssen, personal fees from Servier, personal fees from Medice, outside the submitted work.

ADK reports grants and personal fees from Janssen Pharmaceuticals, grants from Axsome Pharmaceutics, grants from Reveal Biosensors, grants from The Ray and Dagmar Dolby Family Fund, grants from National Institutes of Health, personal fees from Adare, personal fees from Axsome Therapeutics, personal fees from Big Health, personal fees from Eisai, personal fees from Evecxia, personal fees from Ferring Pharmaceuticals, personal fees from Galderma, personal fees from Harmony Biosciences, personal fees from Idorsia, personal fees from Jazz Pharmaceuticals, personal fees from Millenium Pharmaceuticals, personal fees from Merck, personal fees from Neurocrine Biosciences, personal fees from Neurawell, personal fees from Pernix, personal fees from Otsuka Pharmaceuticals, personal fees from Sage, personal fees from Takeda, personal fees from Harmony, outside the submitted work.
ML has nothing to disclose.

RHM-W reports personal fees from Janssen-Cilag, personal fees from LivaNova, personal fees from Lundbeck, personal fees from P1 Vital Ltd, personal fees from Pfizer, personal fees from Sage Therapeutics, personal fees from Sunovian, non-financial support from MagStim, grants from AstraZeneca, personal fees from My Tomorrows, personal fees from OCM Communications s.n.c, personal fees from Syntropharma, outside the submitted work.

RSM reports grants from CIHR, grants from GACD, grants from Chinese National Natural Research

Foundation, personal fees from Lundbeck, personal fees from Janssen, personal fees from Purdue, personal fees from Pfizer, personal fees from Otsuka, personal fees from Allergan, personal fees from Takeda, personal fees from Neurocrine, personal fees from Sunovion, personal fees from Eisai, personal fees from Minerva, personal fees from Intra-Cellular, personal fees from Abbvie, other from AltMed, outside the submitted work.

AM-L reports personal fees from Agence Nationale de la Recherche, personal fees from American Association for the Advancement of Science, personal fees from Brain Mind Institute, personal fees from Brainsway, personal fees from Catania International Summer School of Neuroscience (CISSN), personal fees from Daimler und Benz Stiftung, personal fees from Fondation FondaMental, personal fees from Janssen-Cilag GmbH, personal fees from Lundbeck A/S, personal fees from Lundbeck Int. Neuroscience Foundation, personal fees from MedinCell, personal fees from Sage Therapeutics, personal fees from Techspert.io, personal fees from Thieme Verlag, personal fees from von Behring Röntgen Stiftung, personal fees from BAG Psychiatrie Oberbayern, personal fees from Biotest AG, personal fees from Forum Werkstatt Karlsruhe, personal fees from International Society of Psychiatric Genetics, personal fees from Brentwood, personal fees from Klinik für Psychiatrie und Psychotherapie Ingolstadt, personal fees from Lundbeck SAS France, personal fees from med Update GmbH, personal fees from Merz-Stiftung, personal fees from Siemens Healthineers, outside the submitted work.

AHM reports personal fees from Boehringer Ingelheim, outside the submitted work.

CBN reports grants from National Institutes of Health (NIH), personal fees from ANeuroTech (division of Anima BV), personal fees from Taisho Pharmaceutical, Inc, personal fees from Takeda, personal fees from Signant Health, personal fees from Sunovion Pharmaceuticals, Inc, personal fees from Janssen Research & Development LLC, personal fees from Magstim, Inc, personal fees from Navitor Pharmaceuticals, Inc, personal fees from Intra-Cellular Therapies, Inc, personal fees and other from EMA Wellness, personal fees from Acadia Pharmaceuticals, personal fees from Axsome, personal fees from Sage, personal fees from BioXcel Therapeutics, personal fees from Silo Pharma, personal fees from XW Pharma, personal fees from Neuritek, personal fees from Engrail Therapeutics, personal fees from Brain and Behavior Research Foundation (BBRF), personal fees and other from Anxiety and Depression Association of America (ADAA), personal fees from Skyland Trail, personal fees from Signant Health, personal fees from Laureate Institute for Brain Research (LIBR), Inc, personal fees from Magnolia CNS, other from Gratitude America, other from Xhale Smart, Inc, other from Xhale, other from Seattle Genetics, other from Antares, other from BI Gen Holdings, Inc, other from Corcept Therapeutics Pharmaceuticals Company, from null, outside the submitted work; In addition, CBN has a patent Method and devices for transdermal delivery of lithium (US 6, 375, 990B1) issued, and a patent Method of assessing antidepressant drug therapy via transport inhibition of monoamine neurotransmitters by ex vivo assay (US 7, 148, 027B2) issued.

CN reports personal fees from Johnson/Janssen-Cilag, outside the submitted work.

DN reports personal fees from British National Formulary, personal fees from Ranvier, personal fees from Opiant, personal fees from COMPASS Pathways, personal fees from AWAKN, personal fees from Psyched Wellness, personal fees from Lundbeck, personal fees from BMS/Otsuka, personal fees from Janssen, personal fees from Takeda, grants from Wellcome Trust, MRC, other from P1vital, other from Awakn, other from Psyched Wellness, other from Equasy Enterprises, other from Alcarelle, outside the submitted work; and Expert witness related to psychotropic drugs in several legal cases. Edited/written more than 34 books, some purchased by pharmaceutical companies.

SP reports grants from R21 NIMH (R21DA042271-01), outside the submitted work.

LP reports personal fees from University of Modena and Reggio Emilia, personal fees from University of Miami, personal fees from EDRA-LSWR Publishing Company, personal fees from Inpeco SA Total Lab Automation Company, personal fees from VeraSci, Durham, USA, personal fees from AbbVie USA, personal fees from Acadia USA, personal fees from BCG Switzerland, personal fees from Boehringer Ingelheim International GmbH, personal fees from Compass Pathways, personal fees from Ferrer Spain, personal fees from Gedeon-Richter, Hungary, personal fees from Johnson & Johnson USA, personal fees from NeuroCog Trials USA, personal fees from Novartis-Gene Therapies, Switzerland, personal fees from Otsuka USA, personal fees from Pfizer Global USA, personal fees from PharmaMar Spain, personal fees from Relmada USA, personal fees from Takeda USA, personal fees from Vifor Switzerland, outside the submitted work.

BWJHP reports grants from Boehringer Ingelheim, grants from Janssen Research, outside the submitted work.

AFS reports personal fees from Axsome, personal fees and other from Delpor, personal fees and other from NeuraWell, personal fees from ANeuroTech, personal fees from Signant, personal fees from Otsuka, personal fees from EMA Wellness, personal fees and other from Owl Insights, personal fees and other from Alto, personal fees and other from Verso, personal fees from Compass, personal fees from Janssen, personal fees from Tris, personal fees from Schwabe, personal fees from Boehringer Ingelheim, personal fees from McKinsey, other from Corcept, other from Merck, other from Seattle Genetics, other from XHale, other from Intersect, other from Epiodyne, outside the submitted work.

RCS reports grants from Agency for Healthcare Research and Quality, grants and personal fees from Allergan, grants from Assurex Health, grants from Avanir Pharmaceuticals, grants and personal fees from Cerecor, Inc, personal fees from Clintara LLC, grants from Genomind, grants and personal fees from Janssen Pharmaceutica, personal fees from Medtronic, Inc., grants from Novartis, Inc., personal fees from Pfizer, Inc., grants and personal fees from Otsuka Pharmaceuticals, grants and personal fees from Acadia Pharmaceuticals, grants from Alkermes, PLC, grants and personal fees from Takeda Pharmaceuticals, grants from NeuroRx Inc., outside the submitted work.

LNY reports personal fees from Alkermes, grants and personal fees from Allergan, personal fees from CANMAT, grants and personal fees from CIHR, grants and personal fees from Dainippon Sumitomo, grants and personal fees from Intracellular Therapies, grants and personal fees from Lundbeck, personal fees from Merck, personal fees from Otsuka, personal fees from Sanofi, personal fees from Sunovion, outside the submitted work.

AHY reports personal fees from King's College London, personal fees from Sumitomo Dainippon Pharma, grants from MRC (UK), grants from Wellcome Trust (UK), grants from NIHR (UK), outside the submitted work; and Principal Investigator in the Restore-Life VNS registry study funded by LivaNova. Principal Investigator on ESKETINTRD3004: “An Open-label, Long-term, Safety and Efficacy Study of Intranasal Esketamine in Treatment-resistant Depression.” Principal Investigator on “The Effects of Psilocybin on Cognitive Function in Healthy Participants.” Principal Investigator on “The Safety and Efficacy of Psilocybin in Participants with Treatment-Resistant Depression (P-TRD)”.

RZ reports personal fees from The London Depression Institute, personal fees from Lundbeck, personal fees from Janssen, personal fees from Depsee Ltd, personal fees from Scients Institute, Palo Alto, outside the submitted work; and Co-Investigator on a LivaNova Funded observational study of Vagus Nerve Stimulation for Depression. Collaborated with EMIS PLC for a study. Affiliated with the D'Or Institute of Research and Education, Rio de Janeiro.

GA reports personal fees from BfArM (Federal Institute for Drugs and Medical Devices), personal fees from EMA (No money, honoraria or other kind of compensation paid for services rendered. Only travel and accommodation expenses are covered together with travel allowance from EMA), personal fees from University of Thessaly (Greece) and University of Nicosia (Cyprus) (Honoraria for lectures in Postgraduate Program) outside the submitted work.

FB-D has nothing to disclose.

CF reports personal fees and other from GSK, outside the submitted work.

MH reports personal fees from BfArM, personal fees from EMA (No money, honoraria or other kind of compensation paid for services rendered. Only travel and accommodation expenses are covered together with travel allowance from EMA), personal fees from Forum-Institut, outside the submitted work.

TL reports personal fees from Massachusetts General Hospital Clinical Trials Network and Institute, personal fees from Acadia, personal fees from Alairion, personal fees from Antares, personal fees from Aptinyx, personal fees from Arbor, personal fees from AxsomeBetterLife, personal fees from Biohaven, personal fees from Cadent, personal fees from Cereval, personal fees from Cingulate, personal fees from Corteyxme, personal fees from Eleusis, personal fees from Emerald Lake Safety, personal fees from Greenwich, personal fees from Harmony, personal fees from Janssen, personal

fees from Karuna, personal fees from KemPharm, personal fees from LB Pharmaceuticals, personal fees from LevoTx, personal fees from MAPS, personal fees from Merck, personal fees from Minerva, personal fees from Neurocrine Biosciences, personal fees from Neuronetics, personal fees from Noema, personal fees from Novartis Pharma AG, personal fees from Novartis US, personal fees from Praxis Biosciences, personal fees from Promentis, personal fees from Tonix, personal fees from Vallon, personal fees from Waypoint, outside the submitted work.

F-LM reports personal fees from HUS Helsinki University Hospital, non-financial support from European Patients' Forum, outside the submitted work.

KS has nothing to disclose.

AT has nothing to disclose.

GA-H reports grants from European Commission (EU-PEARL, IMI2 Grant Agreement no. 853966), during the conduct of the study; grants from European Union's Horizon 2020 Research and Innovation Program (grant agreement no.728018), outside the submitted work.

FB reports grants from European Commission (EU-PEARL, IMI2 Grant Agreement no. 853966), during the conduct of the study.

SMG reports grants from European Commission (EU-PEARL, IMI2 Grant Agreement no. 853966), during the conduct of the study; personal fees from Almirall SA, personal fees from Celgene, personal fees from Forum für Medizinische Fortbildung (FomF), personal fees from Mylan GmbH, non-financial support from GAIA Group, grants from Deutsche Forschungsgemeinschaft, grants from German Federal Ministry of Health, grants from National Multiple Sclerosis Society, outside the submitted work.

WJGH reports grants from European Commission (EU-PEARL, IMI2 Grant Agreement no. 853966), during the conduct of the study.

EM reports grants from European Commission (EU-PEARL, IMI2 Grant Agreement no. 853966), during the conduct of the study; personal fees from Lundbeck, personal fees from Janssen-Cilag, outside the submitted work.

JM reports non-financial support from EU-PEARL Project, during the conduct of the study; personal fees from Janssen, outside the submitted work.

CO reports grants from European Commission (EU-PEARL, IMI2 Grant Agreement no. 853966), during the conduct of the study; personal fees from Ferring, personal fees from Janssen, personal fees from Lundbeck, personal fees from SAGE Therapeutics, personal fees from Fortbildungskolleg,

personal fees from Limes Klinikgruppe, personal fees from Medical Tribune, grants from German Research Foundation (OT 209/7-3; 14-1, EXC 2049), grants from German Federal Ministry of Education and Research (KS2017-067), grants from Berlin Institute of Health (B3010350), outside the submitted work.

MES reports non-financial support from EU-PEARL Project, during the conduct of the study; personal fees from Janssen Pharmaceutica NV, other from Johnson & Johnson, other from Novartis Inc., outside the submitted work.

EvdK reports grants from European Commission (EU-PEARL, IMI2 Grant Agreement no. 853966), during the conduct of the study.

KW reports non-financial support from EU-PEARL Project, during the conduct of the study; personal fees from Janssen R&D LLC, other from Johnson & Johnson, outside the submitted work.

YF reports non-financial support from EU-PEARL Project, during the conduct of the study; personal fees from Novartis Pharma AG, outside the submitted work.

JAR-Q reports grants from European Commission (EU-PEARL, IMI2 Grant Agreement no. 853966), during the conduct of the study; personal fees and non-financial support from Eli-Lilly, personal fees and non-financial support from Janssen-Cilag, personal fees from Novartis, personal fees and non-financial support from Shire, personal fees and non-financial support from Takeda, personal fees from Bial, personal fees and non-financial support from Shionogi, personal fees from Lundbeck, personal fees from Almirall, personal fees from Braingaze, personal fees from Sincrolab, personal fees and non-financial support from Medice, personal fees and non-financial support from Rubió, personal fees from Raffo, personal fees from Actelion, personal fees from Ferrer, personal fees from Oryzon, personal fees from Roche, personal fees from Psious, outside the submitted work.

AJS reports non-financial support from EU-PEARL Project, during the conduct of the study; personal fees from Johnson & Johnson, outside the submitted work (employee and own stock in the company).

CMP reports grants from European Commission (EU-PEARL, IMI2 Grant Agreement no. 853966), during the conduct of the study. Additional funding has been provided by UK National Institute of Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Trust and King's College London. He is also funded by a NIHR Senior Investigator Award. He has received research or consultation funding from Johnson & Johnson and Boehringer Ingelheim, as well as research funding from the Medical Research Council (UK) and the Wellcome Trust for research on depression and inflammation as part of two large consortia that also include Johnson & Johnson, GSK and Lundbeck.