Appendix B: Protocol for DSM-5 Field Trials in Academic/Large Clinic Settings

(Revised 4/28/10. Revisions highlighted in red.)

1.1 Specific Aims:

The overall goal of the DSM-5 revision process is to publish a DSM that helps clinicians to provide patients with the best care possible and, at the same time, provides researchers with a tool that will enable continued advancement of the science of psychiatry. It is therefore important to ensure that the proposed diagnostic criteria (and cross-cutting dimensional measures and diagnostic-specific severity measures) are subjected to rigorous and empirically-sound field trials (FTs) before DSM is released for general clinical and research use. Specifically, FTs that test the clinical utility, feasibility, reliability, and, where possible, convergent validity of the proposed diagnostic criteria and dimensional measures, in the environments in which they will be used, are critical. As such, the DSM-5 FTs will aim to address revisions of criteria and additions of severity and cross-cutting measures by posing the following questions:

Questions related to the Diagnostic Criteria:

a. Are the proposed diagnostic criteria and the associated diagnostic checklist easy to understand and use?
b. Do the proposed diagnostic criteria accurately reflect or capture their patients' symptom presentations?
c. Are the proposed diagnostic criteria and the associated diagnostic checklists useful/helpful in formulation of treatment plans for patients?
d. Are the proposed diagnostic criteria reliable in capturing the patients' symptom presentation?
e. Is there convergent validity of diagnoses from expert panel reviews (of videotaped interviews)?

Questions related to the Diagnostic-specific Severity Measures

f. Are the diagnostic-specific severity measures easy to understand and incorporate into the clinical evaluation of patients?
g. Are the diagnostic-specific severity measures reliable?
h. Are the diagnostic-specific measures able to capture change in the severity of symptoms over time (*i.e., sensitivity/responsiveness to change*)?

Questions related to the Cross-Cutting Dimensional Measures

i. Do patients find the cross-cutting dimensional measures easy to use?
j. Are the cross-cutting dimensional measures easy to understand and incorporate into the clinical evaluation of patients?
k. Are the cross-cutting dimensional measures reliable?
l. Are the cross-cutting dimensional measures able to capture change in the severity of symptoms over time (*i.e., sensitivity/responsiveness to change*)?

1.2 Field Trial Settings:

Since DSM is primarily a clinical tool, it is important that we answer these questions in clinical populations, such as diagnosis-specific mental health clinics, general psychiatry clinics, general medical clinics (i.e., primary care), and solo and small group practices, since they represent the bulk of settings in which individuals seek care for their mental health problems. FTs in a variety of clinical settings are necessary given that feasibility, clinical utility, reliability, and validity of a measure (or diagnostic criteria) may differ from one clinical population to another. In addition, since DSM is used across a number of disciplines, it is also necessary that FTs reflect this variation.

The task of conducting DSM-5 FTs in various clinical settings and disciplines poses many challenges for its design. Factors that affect the design include differences in patient volume and clinician staffing and in the prevalence and types of mental disorders seen across settings. Of equal importance are the differences in practice infrastructures, which impact what can feasibly be examined in the different settings. Although it is possible to examine feasibility, clinical utility, and sensitivity to change of the proposed diagnostic criteria and dimensional measures across all settings, it may be difficult to examine reliability and some measures of convergent or criterion validity in solo practitioners' offices. The absence of a trained co-rater would preclude reliability assessments, and there may be problems with obtaining informed consent in this context for procedures (e.g., video-taping) that deviate from usual practice.

FTs to examine reliability, sensitivity to change, and validity are more feasible in some general psychiatric and specialty mental health settings with multiple clinicians and a research infrastructure in place. Even there, videotaping interviews for subsequent inter-rater reliability examination or to establish a criterion validity standard can pose both logistical and IRB issues. Primary care settings pose additional challenges for FTs in that, not only is the prevalence of mental disorders lower in such settings, but primary care providers typically diagnose mental disorders at a more general level compared to psychiatrists, have less familiarity with mental disorders and their diagnostic criteria, have high patient volume but limited time for patient assessment, and have longer patient follow-up gaps. These factors significantly affect what can be examined in FTs involving primary care settings (i.e., which diagnostic criteria and whether testing of reliability and validity is possible). In short, every setting in which FTs might be carried out poses special challenges that need to be addressed carefully.

1.3 Study Designs:

Two different study designs are proposed for the DSM-5 FTs. These designs are related to the settings in which they will be conducted. One design allows for the demonstration of clinical utility, feasibility, reliability, validity, and sensitivity to change while the other focuses on only clinical utility, feasibility, and sensitivity to change. Each design, along with the measures that will be used in the FTs, will be pilot tested and feasibility assessed across a few settings. The results from the pilot testing will inform relevant changes prior to the start of the FTs.

Briefly, the FT design for the academic/large clinic settings is as follows:

1.3.1 ***Stratified Sampling Design for FTs in Academic/Large Clinic Settings:***

a. Overview:

The study will be primarily cross-sectional in nature but will also have a longitudinal component involving 2 follow-up visits. The first follow-up visit will occur at least 4 hours but no more than 2 weeks after the baseline study visit. The second follow-up visit will occur 4-12 weeks after the baseline visit. The first follow-up visit is included to enable the examination of the test-retest reliability of the proposed DSM-5 diagnostic criteria, the diagnostic-specific severity measures and the cross-cutting dimensional measures. The 4-12 week follow-up visit is included to allow for the examination of the ability of the cross-cutting dimensional measures and the diagnostic-specific severity measures to capture change in patients' symptomatology over time (*i.e., responsiveness to change*).

b. Sampling Strategy

For the DSM-5 Field Trials in Academic or Large Clinical settings, a *stratified sample (SS) design* will be used. This will involve the sampling of consecutive patients entering a particular clinical setting for their initial or routine clinic visits during the 4½ month recruitment period for the DSM-5 FT for entry into the FT. The pool of consecutive patients includes new patients to the setting who are assigned their treating clinicians from among the pool of study clinicians (i.e., an assigned treating clinician [ATC]) within the setting as well as those who are existing patients in the setting with an existing treating clinician (ETC). The pool of study clinicians from which each NP is assigned a treating clinician will exclude DSM-5 Task Force and/or Work Groups members or advisors to the DSM-5 process working at the site. As outlined in **Figure 1**, consecutive new and existing patients will see their treating clinician for their respective initial or routine clinic visits that occur during the DSM-5 FT period. The treating clinician will carry out this patient visit as clinically indicated for the respective patient. The treating clinician is asked to complete a brief Patient Recruitment/Screening form for each patient. The Patient Recruitment/ Screening Form inquires about the patient's diagnoses, whether the patient is currently symptomatic, has past or current self-injurious ideation or behaviors, and other age-specific and relevant factors. The treating clinicians will inform their currently symptomatic patients about the DSM-5 FT and asked if they are interested in participating in the study. *This pool of consecutive new and existing patients who are symptomatic at the time of their initial or routine clinic visits represents the sampling frame for the FT at the particular site.*

Treating clinicians will ask their patients who expressed interest in participating in the DSM-5 FT to the meet with the site research coordinator for 10-15 minutes to determine their eligibility. The site research coordinator will use the Patient Recruitment/Screening Form to help determine the patient's eligibility. Eligible patients will be asked to provide written consent. Eligibility criteria are based on whether patients are able to read and communicate in English (unless patients, interview forms and clinicians use another language) and whether they have the DSM-IV diagnoses of interest for the particular FT site (as determined by each patient's treating clinician and documented on the Patient Recruitment/ Screening Form). **For disorders that are new to DSM-5**, a computer algorithm [available in the electronic data collection system and to the site research coordinator] will be used to

determine diagnostic eligibility based on the presence of relevant DSM-IV diagnoses and specific symptomatology documented by the patient's treating clinician on the Patient Recruitment/Screening Form (e.g., presence of temper tantrums for Temper Dysregulation Disorder with Dysphoria, past or current self-injurious ideation or behaviors for Non-suicidal Self Injury). The computer algorithm will require the site research coordinator to simply enter the information included in the Patient Recruitment/ Screening Form.

During the consent process the research coordinator will inform each patient that by agreeing to participate in the study he/she will be asked to return to the clinic for 3 additional visits (i.e., the baseline, a follow-up visit that can occur no less than 4 hours but no greater than 2 weeks following the baseline visit, and a second follow-up study visit at 4-12 weeks following the baseline study visit). Patients will also be informed that one (for new patients) or two (for existing patients) of the clinical interviews will be conducted by a new clinician for a second opinion on his/her diagnosis while the treating clinician observes and does his/her own ratings. In addition, each patient will be informed of the possibility of being selected to have his/her baseline study interview videotaped for a review of his/her diagnoses by an expert panel, if the site is doing videotaping. The site research coordinator will collect information on the proportion of patients seen in the specific site with the disorders of interest over the course of the study. This information will be used to create sampling weights (i.e., sampling weights based on pre-selected sample), which will be used across all analyses to account for the sampling scheme of the study and provide estimates that are representative of the target population.

For each FT site, the site research coordinator will select each eligible and consenting new and existing patient into a diagnostic group (i.e., D) or a control group (i.e., Other) based on the disorders being tested at the site. As indicated earlier, for the testing of disorders that are new to DSM-5 a computer algorithm will be used to determine the patient's diagnostic eligibility and the group to which the patient should be assigned. There will be a minimum of 2 to a maximum of 5 diagnostic groups (i.e., D2-D5) and one control group (i.e., Other) at any one site. Each diagnostic group will consist of 50 new and existing patients combined while the control group will comprise 50 randomly selected patients from among the pool of new and existing patients without any of the diagnoses of interest. Each selected new and existing patient is scheduled to return to the clinic within two weeks for the **baseline study visit**. The discrepancies between i) those who are informed of the FT and who expressed interest and met with the site research coordinator, ii) those who met with the site research coordinator and those who provided written consent, and 3) those who provided written consent and those who return to the clinic for the baseline study visit will be documented.

1.3.2 **Study Visits:**

The DSM-5 FT will involve 3 study visits: a baseline study visit and 2 follow-up visits.

a. **The Baseline Study Visit (Figure 1a):**

The **baseline study visit** involves each new and existing patient within each diagnostic group being asked to fill out the cross-cutting measures (level 1 and indicated level 2). Following the completion of the cross-cutting measures, a new patient sees his assigned treating clinician for completion of the baseline clinical interview. However,

after completing the cross-cutting measures, an existing patient sees a new clinician (**Clinician 1 [C1]**) who is sampled from the pool of participating study clinicians at that setting excluding his/her existing treating clinician and any members of the DSM-5 Task Force and/or Work Groups or advisors to the DSM-5 process at the site. A clinician who is new to the participating patient (**C1**) is used for the initial evaluation since a primary goal of the FT is to test the reliability of the proposed diagnostic criteria and the cross-cutting and diagnostic-specific severity measures.

Baseline Interview Logistics:
The use of a clinician other than the patient's existing treating clinician (i.e., C1) to test reliability ensures independence of errors and an unbiased estimation of the reliability coefficient. In other words, given that the study does not propose the use of a structured interview, if the routine clinician for an existing patient, especially with long-term contact, is used to conduct the clinical interview, his or her prior knowledge of the patient will affect the questions asked and probes used throughout the interview and subsequently his or her diagnosis of the patient. Such biased probing and questioning will impact the observing clinician's diagnostic formulation of the patient, thereby inflating the reliability coefficient obtained. However, without a structured diagnostic interview, the reliability assessment of the diagnostic criteria becomes quite challenging in that clinicians assessing the same patients within a short time span for test-retest will have minimal or no guidance on how to operationalize the diagnostic criteria, which can lead to variability in the diagnoses made, and impact the reliability coefficient obtained. Therefore, as with all reliability studies, the sample size determination is important. The sample size for the study (i.e., 50 per diagnostic group for this study) needs to be sufficiently large to have the power to account for the various sources of variability, such as by chance, across clinicians and across patients within the same diagnostic groups. This same argument applies if the same interviewer is used to conduct the baseline clinical evaluation and diagnostic formulation and the clinical evaluation and diagnostic formulation that occurs at the first follow-up visit, since his or her questions and probes will be influenced by his or her prior assessment of the patient. Hence, different clinicians will be needed for test and retest.

Use of a clinician other than the treating clinician for the clinical evaluation poses its own challenges for continuity of care for the patient. As such, the treating clinician will observe the clinical interview for visits that involve the use of a new clinician in the determination of reliability. In such instances, the treating clinician will complete his or her own diagnostic formulation using the diagnostic checklist and any indicated dimensional measures. If the treating clinician can provide a longitudinal examination of all the data (LEAD standard) for the patient, there will also be an opportunity to compare this with a cross-sectional diagnosis by the independent clinicians.

The study clinicians at each study site will be informed of the disorders of interest at the particular site as well as of the fact that the study involves a control group that consists of a random sample of patients without any of the diagnoses of interest.

However, each clinician will be blind to each patient's group assignment. The clinicians are asked to conduct the clinical evaluation for each study patient the way they would for any patient coming to see them for the first time and to diagnose primary and any comorbid conditions that the patient may have. The clinician has the option to select any DSM-5 Diagnostic Checklist that he/she deems appropriate based on the patients presenting symptoms as determined from the clinical evaluation. The DSM-5 Diagnostic Checklists will be available in the Electronic Data Collection system being used for the DSM-5 Field Trials (i.e., REDCap developed by Vanderbilt University through funding from NIH-DRR (1) and customized for the DSM-5 Field Trials). The REDCap system will allow two clinicians who are rating and completing clinician-rated scales and DSM-5 Diagnostic Checklists for the same patient to independently log in to the system, access the results of the patient-rated assessment, and access and complete the clinician-rated scales and DSM-5 Diagnostic Checklists that he/she deems appropriate, while being blind to each other's ratings.

The Baseline Clinical Interview:

As part of clinical interview at the **baseline study visit**, the assigned treating clinician for a new patient and C1 for an existing patient will examine the patient's self-rated cross-cutting measures (and, if necessary, complete any clinician-rated level 2 cross-cutting measures), conduct his/her clinical evaluation of the patient, formulate one or more diagnoses, complete the relevant DSM-5 Diagnostic Checklist/s of draft diagnostic criteria, complete any indicated diagnostic-specific severity measures, and formulate a treatment plan for the patient. For an existing patient, the existing treating clinician will observe the interview, review the patient's self-rated forms, complete his or her own ratings, formulate one or more diagnoses, complete the relevant DSM-5 Diagnostic Checklist/s of draft diagnostic criteria, write a chart note, and discharge the patient in private with his/her treatment plan and needed prescriptions. All clinicians who participate in the baseline study visit for a patient will complete the Clinical Utility Questionnaire at the end of the visit. In addition, for cases where the treating clinician observed the clinical interview and completed his/her own rating and diagnosis of the patient, he/she will complete questions about the clinical history and current mental status information obtained at this single interview by C1. Once the independent ratings are recorded, discrepancies between raters will be identified and both clinicians will identify perceived reasons for any diagnostic or severity rating discrepancies using the Discrepancy Interview Protocol methodology [developed at Washington University (2)].

Videotaping Option:

In sites where the technical capability exists, the baseline clinical interview will be videotaped for a random 20% of new and existing patients combined. These tapes will be used for review of the diagnoses by a panel of experts (i.e., the validity component of the FT) as well as the assessment of inter-rater reliability of the diagnostic criteria and the cross-cutting and diagnostic-specific severity measures. This random 20% represents 10 patients per selected diagnostic group for a maximum of 100 videotapes. The validity component of the FT, which will be conducted on selected diagnostic

criteria, will involve the review of each videotape and the associated patient self-rated forms by an expert panel comprising three senior clinicians at the specific clinical site. The senior clinicians will include members of the DSM-5 Task Force or Work Groups or advisors to the DSM-5 process who are located at the particular site.

For selected patients, videotaped interviews may be recommended for the development of case studies. To protect patient privacy, permission would be requested for actors to recreate patients' responses in a simulated interview that could be used for teaching purposes. Such videotaped, simulated case studies that include multiple diagnoses and controls with sub-threshold or no diagnoses may be used for additional reliability and validity studies.

b. **The first follow-up visit (Figure 1b)***:*

At the end of the baseline study visit, each selected patient within each diagnostic group is scheduled for a follow-up study visit at least 4 hours but no later than 2 weeks following the baseline clinical interview. At this **follow-up visit**, each patient completes the self-rated level 1 and indicated level 2 cross-cutting measures, and will then see an independent clinician (i.e., **C1** for new patients and **Clinician 2** [C2] for existing patients) for the study's second clinical interview. The new clinicians will come from the same sample of study clinicians as the assigned treating clinician for new patients and C1 for existing patients during the baseline visit as outlined above.

Second Clinical Interview:

The study's second clinical interview will be conducted by **C1** for new patients and **C2** for existing patients. However, the interview will be observed by the patient's treating clinician. The new clinician will be blinded to all information obtained about the patient in the baseline interview as well as the patient's group assignment. In addition, the new clinician and the treating clinician who observes the clinical interview will independently review the self-rated forms completed by the patient at the start of this visit, complete any indicated clinician-administered level 2 cross-cutting measures, formulate their diagnoses of the patient based on the clinical interview and complete the relevant DSM-5 Diagnostic Checklists, complete any indicated diagnostic-specific severity measures while blinded to each other's ratings. At the end of this study visit, the treating clinician will write a chart note and privately discharge the patient with his/her treatment plan and needed prescriptions. Both clinicians who participate in this follow-up visit for each patient will complete the Clinical Utility Questionnaire at the end of the visit. In addition, the treating clinician will complete questions about the adequacy of the clinical history and current mental status information obtained at this single interview conducted by the new clinician, and both clinicians will provide reasons for any discrepancies between their ratings and diagnoses using the Discrepancy Interview Protocol (2). The treating clinician will also provide reasons for any discrepancies between his/her ratings and diagnoses at baseline and the first follow-up visits follow-up using the Discrepancy Interview Protocol (2). Throughout this process, the treating physician should not discuss the patient with either C1 or C2.

c. **The Second Follow-up Visit (Figure 1c):**

At the end of the first follow-up study visit, each selected patient within each diagnostic group will be scheduled for a second follow-up study visit, which will occur at 4-12 weeks after the baseline assessment. At this **4-12 week follow-up visit**, each patient completes the self-rated level 1 and indicated level 2 cross-cutting measures and is then seen by his or her treating clinician.

Third Clinical Interview:

For each selected patient in each diagnostic group, the study's third clinical interview will be conducted by his/her treating clinician. The treating clinician will review the patient's diagnoses and complete any indicated clinician-administered level 2 cross-cutting and/or diagnostic-specific severity measures. The treating clinicians' assessments at the baseline and the two follow-up visits will be used to assess the responsiveness to change of the clinician-completed cross-cutting and diagnostic-specific severity measures.

1.4 Validation of Select DSM-5 Diagnostic Criteria:

The videotapes of the 20% randomly selected baseline clinical interviews **within specific diagnostic groups** will be used for validation of DSM-5 diagnostic criteria for the relevant disorders. Each videotape will be reviewed by a 3 member panel of experts for the determination of consensus diagnosis. Each panel of experts will include members of the DSM-5 Task Force and/or Work Groups or advisors to the DSM-5 process who are located at the clinical site. For the validation of the diagnostic criteria for a particular disorder the following steps will be taken:

1. Each member of the expert panel will:
    a. View each of the videotapes associated with the particular diagnostic criteria at that site
    b. Review the associated patient's self-rated and clinician-administered cross-cutting dimensional measures and diagnostic-specific severity measures that were collected at the baseline study visit.
    c. Use the available information to formulate his/her diagnoses, for the patient, using the DSM-5 Diagnostic Checklist
    d. Provide the reasons for his/her decisions.

2. Each expert will be blind to the DSM-V diagnosis assigned by the study clinicians.

3. The final expert diagnosis will be based on consensus as determined by agreement between at least 2 experts.

4. In cases where all three experts disagree, the panel will be brought together to discuss the disagreements and identify reasons for the discrepancies.

1.5 Background Information for Data Analysis:

The analytic strategy will be presented on a broad category basis. All analyses will be conducted using SAS statistical software and the sampling weights. In addition, SAS macro (%QLS) will be used to conduct the quasi-least squares (QLS) computational approach for estimation of the correlation parameter in GEE analyses for data with multiple sources of correlation.

a. *Feasibility:* Feasibility will be assessed by examination of clinician and patient participation rates, overall and item response rates, and the amount of time taken by the patient or clinician to complete the battery of tests. Analysis will be primarily descriptive, e.g., frequency distributions and measures of central tendency (means and their associated standard deviations, modes and median) with some comparisons based on psychiatrists and patients characteristics. Items with greater than a 20% non-response rate will need to be reexamined. The results will be used to further refine the content of the assessment battery, the length of testing, etc. In addition, descriptive analyses will be conducted on clinicians and patients report on the ease of use and understanding of the diagnostic criteria or dimensional measures.

b. *Clinical Utility:* Descriptive statistics will be used to describe the range of subjective and objective responses obtained in the clinical utility component of the study. In addition, common themes will be identified in the open-ended responses and reported. The results of the subjective and objective reports will be compared using appropriate descriptive and multivariate analyses controlling for clinicians' characteristics such as age, sex, and years in practice.

c. *Reliability:* The test-retest and inter-rater reliabilities of the categorical diagnoses and the dimensional measures that are incorporated into the diagnostic scheme for DSM-5 will be assessed in independent analyses. Test-retest and inter-rater reliabilities for dichotomous and categorical items will be computed using the kappa statistic (k) for categorical data; for ordinal and continuous measures, test-retest reliability will be computed using the intraclass correlation coefficient (ICC). In both cases, sampling weights will be incorporated into the computation. The 95% confidence intervals will be obtained using bootstrap methods. Reliability will be rated as fair (.30 to .49), moderate (.50 to .69), or high (.70 to 1.00) for the purposes of comparison (3). Post hoc exploratory analyses will be conducted to determine whether reliabilities differ by gender or due to comorbid conditions.

d. *Sensitivity/Responsiveness to Change:* Mixed models for longitudinal data (linear for continuous outcomes and logistic for categorical outcomes) will be used to examine sensitivity to change for each severity measure with adjustment for age and sex. The mixed model approach is more robust and less sensitive to missing data and variations in follow-up time Variations in follow-up time are likely to occur given the flexibility the clinicians will have in scheduling their patients for the follow-up

assessments. For mixed models, the model building guidelines outlined in Verbeke and Molenberghs (4) will be followed (*i.e., 1. model the structure of means using fixed effects; 2. specify a covariance structure both between subjects and within subjects; 3. fit the means model accounting for the covariance structure specified; and 4. make tests and inferences, including simplifying the means model if possible*). The degree to which participants' ratings of change in severity of symptoms is related to clinicians' ratings of change in severity for the same symptoms is important in determining the potential utility of these patient-rated measures in routine clinical practice. This will be examined using linear regression within a generalized estimating equations (GEE) framework. However, a QLS computational approach for estimation of the correlation parameter will be used, instead of the conventional approach in GEE analyses, given its advantage in handling data with multiple sources of correlation as would be involved with these data (i.e., correlation involved in repeated patient assessment, in repeated clinician assessment, and the impact of clinicians' knowledge of patients' self-rating of symptom severity prior to their assessment of patients' symptom severity).

e. *Validation of Selected Diagnoses:* The proportion of total agreement (between consensus diagnosis by the expert panel and diagnosis by the study clinician), and kappa coefficients with 95% confidence intervals, sensitivity, specificity, and positive and negative predictive values for each condition will be examined. In addition, the concordance between the two diagnostic sources will be assessed using McNemar's test (5). Finally, a series of multivariate logistic regression models will be used to determine factors that predict agreement in diagnostic decisions between experts and the study clinicians. Clinician (e.g., age, sex, profession, and number of years in practice) and patient characteristics (e.g., age, sex, diagnosis, and treatment status) will be examined as potential predictors.

1.6 Sample Size and Power:

Power Analysis and Sample Size software was used to determine the sample size needed to detect a minimum required kappa or ICC value of 0.3 (fair reliability) and an acceptable kappa or ICC value of 0.6 (moderate reliability) with alpha and beta error level of 0.05 and 0.20 respectively given 2 repeated assessments (test-retest reliability) or 2 clinician raters (inter-rater reliability). A sample size of 50 participants is needed within each stratum. This sample size is consistent with that reported by Donner and colleagues in their multiple reports on the sample size requirements for reliability studies (6-9). Using clinically significant change in severity of depression as an example [using 5-point change on the PHQ-9 as an established indicator of clinically significant change (10)], the sample size will also allow us to have sufficient power (beta = 0.20) to detect clinically significant changes in severity even with adjustments for age and sex.

1.7 Dissemination strategy:

Knowledge translation will occur throughout the project. This will include using the information to inform revision of the diagnostic criteria for the disorders relevant to this

study. In addition, the results will be shared with relevant professional and consumer groups for their feedback so as to inform the DSM-5 process. Dissemination of the results of the study will also occur at local, national and international scientific meetings by members of the research team. Reports on the results of this study will also be disseminated in the form of publication in peer-reviewed scientific journals as well as the DSM-5 source books. Following consolidated data analysis for the principal publication, each site will retain the right to publish subsequent analyses of their own data sets. Recognition will be given to all key participants at each site.

1.8 Summary:

This design will accommodate testing of the diagnostic criteria and dimensional measures for mental disorders of high and low prevalence, with prevalence estimates determining the setting in which the FT occurs. Whether for high or low prevalence disorders, patients with the diagnoses of specific interest at their initial or routine clinic visits will be oversampled, and those without them will be under-sampled for the study's baseline and two follow-up visits.

1.9 References:

1. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N and Conde JG Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics, 2009; 42(2):377-381.
2. Cottler LB, Grant BF, Blaine J, Mavreas V, Pull C, Hasin D, et al. Concordance of DSM-IV alcohol and drug use disorder criteria and diagnoses as measured by AUDADIS-ADR, CIDI and SCAN. Drug Alcohol Depend 1997 Sep 25;47(3):195-205.
3. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-74.
4. Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. 2000. New York, NY, Springer-Verlag.
5. McNemar Q: Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947**,** 12:153-157.
6. Donner A, Eliasziw M. Sample size requirements for reliability studies.Stat Med 1987 Jun;6(4):441-8.
7. Eliasziw M, Donner A. A cost-function approach to the design of reliability studies. Stat Med 1987 Sep;6(6):647-55.
8. Shoukri MM, Donner A. Efficiency considerations in the analysis of inter-observer agreement. Biostatistics 2001 Sep;2(3):323-36.
9. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med 1998 Jan 15;17(1):101-10.
10. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002 Mar;64(2):258-266.

Figure 1: Stratified Sampling Design for the DSM-5 Field Trials in Academic/Large Clinical Settings. **(Revised 4/28/10. Revisions highlighted in red.)**

Visit Number	Sampling procedure:
Screen (Each ETC will be informed of the diagnoses of interest) If consent obtained Baseline Visit is scheduled. Each NP/EP is selected into 1 diagnostic group (2-5 groups per site, i.e., D1-D5) or a control group (1 group per site, i.e., Other). There will be a minimum of 150 to a maximum of 300 patients in any one study site.	The sampling frame for the DSM-5 Field Trials (FTs) involves consecutive patients who are presenting at the participating clinical site for their scheduled initial (for new patients) or routine (for existing patients) clinic visits. The patient sample consist of new patients (**NPs**) who are being seen at the site for the first time and existing patients (**EPs**) who already have existing treating clinicians at the site. Each **NP** is randomly assigned to one of the study clinicians (i.e., assigned treating clinician [ATC]) for the initial clinic evaluation. The ATC will carry out the initial clinical evaluation of the NP as is usual for the participating site. For each NP the ATC completes a Patient Recruitment/ Screening Form, which inquires about the patient's DSM-IV diagnoses, whether the patient is currently symptomatic, has past or current self-injurious ideation or behaviors, and other age-specific and relevant factors (such as temper tantrums in children). The ATC informs all symptomatic NPs (regardless of diagnoses) of the DSM-5 FT and asks if he/she would be interested in participating. Each NP is told that he/she may be selected if he/she meets criteria for the FT. Each **EP** sees his/her existing treating clinician (ETC) for his/her routine follow-up appointment. The ETC is asked to document the EP's primary, and if present, any co-morbid DSM-IV diagnoses. For each EP the ETC completes a Patient Recruitment/Screening Form, which inquires about the patient's diagnoses, whether the patient is currently symptomatic, has past or current self-injurious ideation or behaviors, and other age-specific and relevant factors (such as temper tantrums in children). The ETC informs all symptomatic EPs (regardless of diagnoses) of the DSM-5 FT and asks if he/she would be interested in participating. Each EP is told that he/she may be selected if he/she meets criteria for the FT. **NP says NO to FT and is not enrolled.** ATC concludes the patient visit after scheduling any clinically-indicated follow-up appointment. **NPs and EPs who say YES to FT are asked to meet with the Research Coordinator (RC).** RC consents the patient if he/she has any of the diagnoses of interest and informs him/her that he/she may be asked to return to the clinic for 3 additional visits one of which may be videotaped. **EP says NO to FT and is not enrolled.** ETC concludes the patient visit after scheduling any clinically-indicated follow-up appointment. At each study site, each consenting **NP** and **EP** is selected into a diagnostic group (2-5 groups per site, i.e., D1-D5) or a control group (1 group per site, i.e., Other) based on the DSM-IV and symptom information provided by the treating clinician on the Patient Recruitment/Screening Form. Each diagnostic group will consist of 50 NPs and EPs **combined**. The control group will consist of 50 randomly selected patients from among NPs and EPs **without one of the diagnoses of interest** (i.e., without D1, D2, D3, D4, or D5). The number of diagnostic groups will depend on the diagnostic criteria being tested in a particular site**.** The groups **(**i.e., **D1, D2, D3, D4, D5, Other)** will be based on the patient's **DSM-IV diagnosis** as determined by the ATC (for NPs) or the ETC (for EPs). For the testing of disorders that are new to DSM-5, the stratum will be defined by a combination of relevant DSM-IV diagnoses and symptoms as documented by the treating clinician on the Patient Recruitment/Screening Form. The RC will schedule the **BASELINE** study visit for each selected patient. *There will be a minimum of 150 to a maximum of 300 patients in any one study site.*
Baseline visit (**TEST**)[1] *Detailed in Fig 1a*	Each selected patient will go through the procedure for the **BASELINE visit** in which the clinical interview is conducted by the ATC for NPs and a new clinician (C1) for EPs. *For EPs the ETC will observe the clinical interview.* At the end of this visit the first follow-up study visit is scheduled. This first follow-up visit will occur **no less than 4 hours but not greater than 2 weeks** following the baseline visit.
First follow-up visit (**RETEST**)[2] *Detailed in Fig 1b*	Each selected patient will go through the procedure for the **first follow-up study visit** in which the clinical interview is conducted by a new clinician for both NPs (C1) and EPs (C2). *The ATC (for NPs) and ETC (for EPs) will observe the clinical interview.* At the end of the first follow-up visit a 4-12 week follow-up visit is scheduled.
4-12 week follow-up visit (**sensitivity to change**)[3] *Detailed in Fig 1c*	Each selected patient will go through the procedure for the **4-12 week follow-up** visit in which the clinical interview is conducted by the ATC for NPs and ETC for EPs.

[1] For each NP the ATC conducs the interview & formulates diagnoses however for each EP a new clinician (C1) conducts the interview while the ETC observes the interview & does own ratings.
[2] A new clinician (C1 for NPs and C2 for EPs) conducts the interview & formulates diagnoses while the ATC (for each NPs) and the ETC (for each EPs) observes the interview & does own ratings.
[3] The treating clinician (ATC for NPs and ETC for EPs) conducts the interview and completes the same rating forms completed at Baseline.

Figure 1a*: Outline of the steps involved at the **Baseline Study Visit** (TEST)

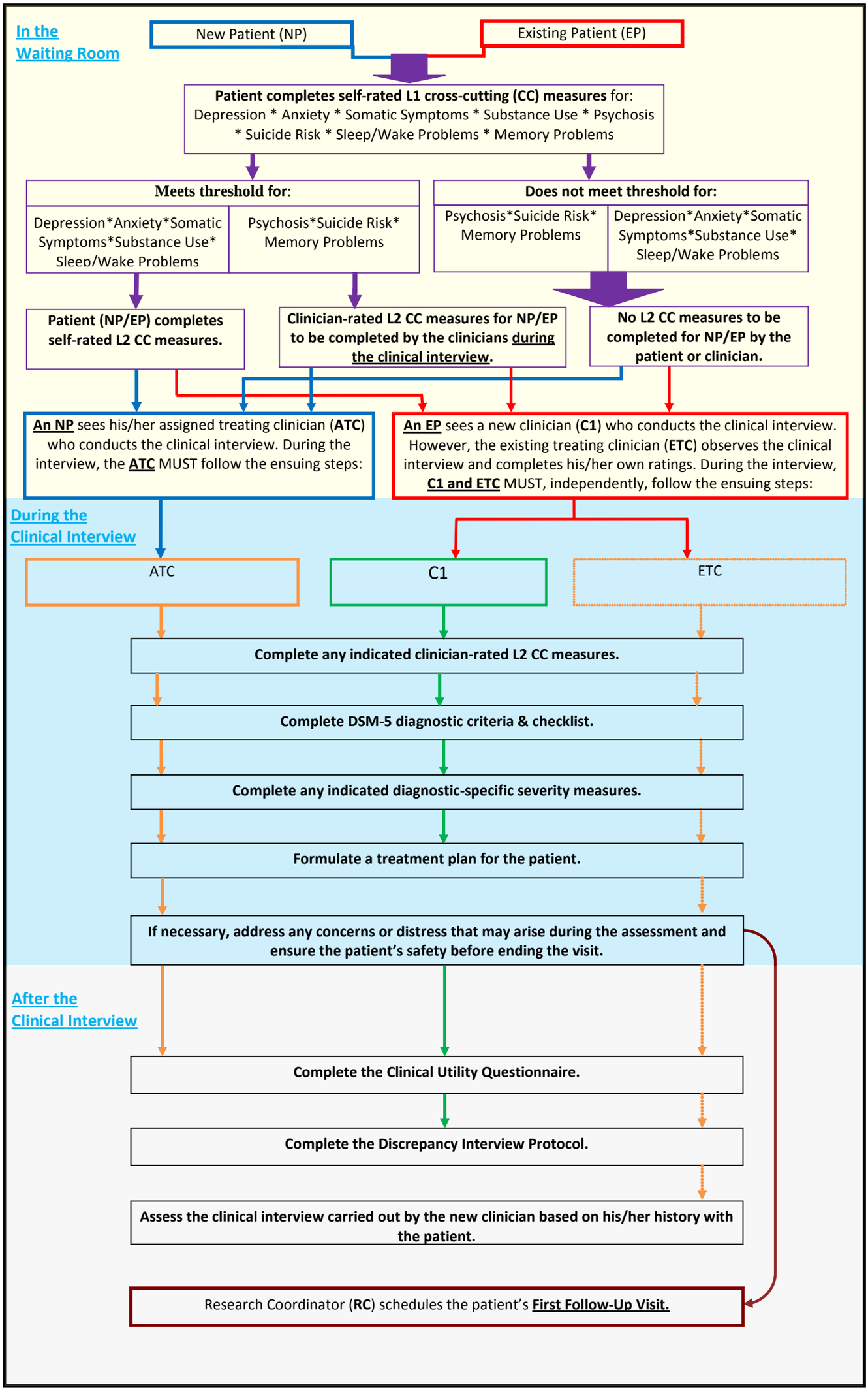


Figure 1b*: Outline of the steps involved at the **First Follow-up Study Visit** (RETEST: 4 hours – 2 weeks after the baseline study visit)

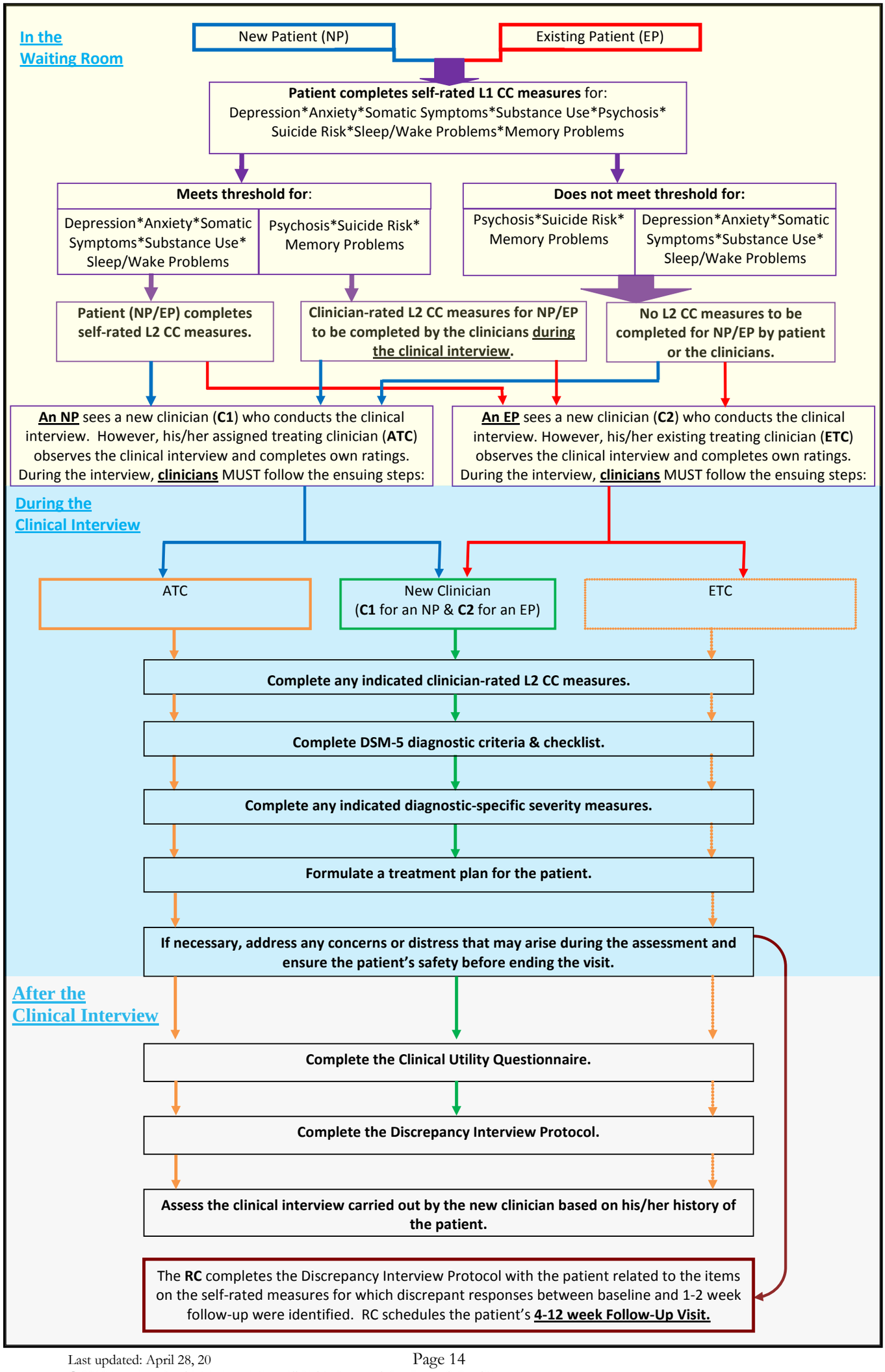
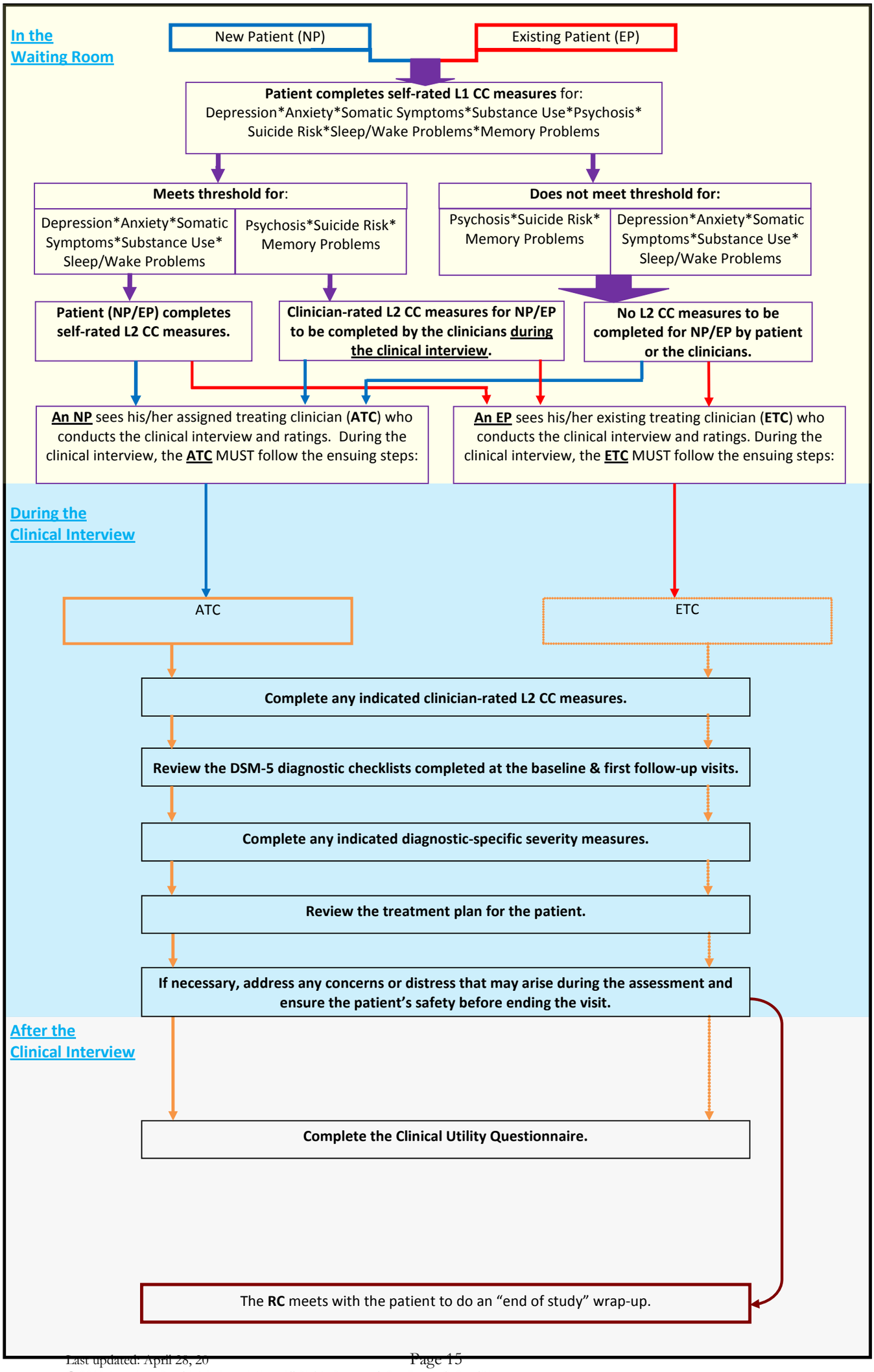


Figure 1c*: Outline of the steps involved at the **4-12 Week Follow-up Study Visit** (Sensitivity to Change relative to Baseline)



*Endnotes:

DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, Version 5.

FT = Field Trials

NP = New Patient. An NP refers to a patient who is new to the clinical setting and must be assigned a treating clinician.
EP = Existing Patient. An EP refers to an existing patient within the clinical setting who has an existing treating clinician.

Please note, NP and EP are used throughout the document for continuity despite the fact that after the baseline visit an NP is no longer new but an existing patient.

ATC = Assigned Treating Clinician. An ATC is a clinician who is assigned as the treating clinician for an NP at the visit prior to the baseline study visit.
ETC = Existing Treating Clinician. An ETC is a clinician who is the existing treating clinician for a patient who was already being seen in the clinical setting.
C1 = New Clinician 1. C1 refers to the first new clinician a patient sees (excluding the ATC)
C2 = New Clinician 2. C2 refers to the second new clinician a patient sees.

Please note, an NP will only see one new clinician (C1) whereas an EP will see two new clinicians (C1 and C2) throughout the study.

CC = Cross-cutting. CC refers to the cross-cutting dimensional measures. These measures have 2 levels:
L1 = level 1 of the cross-cutting measures, which are self- or informant-rated.
L2 = level 2 of the cross-cutting measures. L2 cross-cutting measures can be self-, informant-, or clinician-rated.

Please note, the L2 CC measures that self- or informant-rated will be completed by the patient or informant in the waiting room before going in to see the clinician for the clinical interview. However, the L2 CC measures that are clinician-rated must be completed by the clinician during the clinical interview.

RC = Research Coordinator. Each clinical site involved in the DSM-5 FT in Academic/Large Clinical Settings will have its own RC who will implement the study protocol, recruit patient participants and complete the Discrepancy Interview Protocol with the patients for whom discrepant responses on the self- or informant-rated CC (L1 and L2) are identified when their assessments at baseline and first follow-up visits are compared.