Therapeutic Misconception in Research Subjects: Development and Validation of a Measure

Paul S. Appelbaum, MD, **Milena Anatchkova, PhD**, **Karen Albert, MA**, **Laura B. Dunn, MD**, and **Charles W. Lidz, PhD**

Dr. Appelbaum is Elizabeth K. Dollard Professor of Psychiatry, Medicine, and Law, Department of Psychiatry, Columbia University; Dr. Anatchkova is Assistant Professor of Quantitative Health Sciences, University of Massachusetts Medical School; Ms. Albert is Research Coordinator, Department of Psychiatry, University of Massachusetts Medical School; Dr. Dunn is Associate Professor of Psychiatry, University of California at San Francisco; and Dr. Lidz is Research Professor of Psychiatry, University of Massachusetts Medical School

Abstract

Background—Therapeutic misconception (TM), which occurs when research subjects fail to appreciate the distinction between the imperatives of clinical research and ordinary treatment, may undercut the process of obtaining meaningful consent to clinical research participation. Previous studies have found TM is widespread, but progress in addressing TM has been stymied by the absence of a validated method for assessing its presence.

Purpose—The goal of this study was to develop and validate a theoretically grounded measure of TM, assess its diagnostic accuracy, and test previous findings regarding its prevalence.

Methods—220 participants were recruited from clinical trials at 4 academic medical centers in the U.S. Participants completed a 28-item Likert-type questionnaire to assess the presence of beliefs associated with TM, and a semi-structured TM interview designed to elicit their perceptions of the nature of the clinical trial in which they were participating. Data from the questionnaires were subjected to factor analysis and items with poor factor loadings were excluded. This resulted in a 10-item scale, with 3 strongly correlated factors and excellent internal consistency; the fit indices of the model across 10 training sets were consistent with the original results, suggesting a stable factor solution.

Results—The scale was validated against the TM interview, with significantly higher scores among subjects coded as displaying evidence of TM. ROC analysis based on a 10-fold internal cross-validation yielded AUC=.682 for any evidence of TM. When sensitivity (0.72) and specificity (0.61) were both optimized, Positive Predictive Value was 0.65 and Negative Predictive Value was 0.68, with a Positive Likelihood Ratio of 1.89, and a Negative Likelihood Ratio of 0.47. 50.5% (n=101) of participants manifested evidence of TM on the TM interview, a somewhat lower rate than in most previous studies.

Limitations—The predictive value of the scale compared with the "gold standard" clinical interview is modest, although similar to other instruments based on self-report assessing states of mind rather than discrete symptoms. Thus, although the scale can offer evidence of which subjects are at risk for distortions in their decisions and to what degree, it will not allow researchers to conclude definitively that TM is present in a given subject.

Address correspondence to Dr. Appelbaum at New York State Psychiatric Institute, 1051 Riverside Drive, Unit 122, New York, NY 10032; p: 212-543-4184; f:212-543-6752; psa21@columbia.edu.

Conclusions—The development of a reliable and valid TM scale, even with modest predictive power, should permit investigators in clinical trials to identify subjects with tendencies to misinterpret the nature of the situation and to provide additional information to them. It should also stimulate research on how best to decrease TM and facilitate meaningful informed consent to clinical research.

Keywords

Therapeutic misconception; informed consent; research ethics

Therapeutic misconception (TM) was first described in the 1980s, when it was noticed that some research subjects "fail[ed] to appreciate the distinction between the imperatives of clinical research and of ordinary treatment" [1]. People who manifest TM often express incorrect beliefs about the degree to which their treatment will be individualized to meet their specific needs [2]; the likelihood of benefit from participation in the study [2]; and the goals of the researchers in conducting the project [3]. These beliefs may be attributable to subjects' failure to distinguish their previous experiences receiving medical treatment or to comments made by the research team or on the consent form that foster the conflation of treatment with research [4,5]. The characteristics of TM suggest that it may undercut the process of obtaining meaningful consent to clinical research participation by distorting participants' beliefs about the nature and consequences of the process into which they are entering [4]. In the years since TM was first identified, debates about the precise definition and borders of the concept have multiplied [3],[4],[6], as have discussions of its implications for informed consent to research [4],[7]-[9].

At the same time, a substantial empirical literature has developed, documenting the apparent ubiquity of TM [4]. Among subjects in 44 clinical trials addressing diverse diagnoses, TM was found to be present to some degree in 62% [2]. High TM scores have been identified in 74% of people enrolled in early phase gene transfer trials [5]. Psychiatric research subjects with schizophrenia had manifestations of TM in 69% of cases [10]. TM was found in a pilot study among 12 of 15 Egyptian outpatients [11], and in 70% of people who had consented to research participation for themselves or their children in France [12]. In addition, there are many studies illustrating failures on the part of research subjects to comprehend one or another aspect of clinical research (e.g., unproven experimental treatment, randomization, use of placebo) that seem likely to contribute to TM, e.g., [13]-[16].

Research on TM, however, has been hampered by the absence of both a universally accepted definition and a validated measure of the phenomenon. Scholars have attempted to distinguish TM from related phenomena such as "therapeutic optimism," "therapeutic misestimation," and "unrealistic optimism" [6],[17]. The tendency to apply the term imprecisely, for example, to any belief that research participation could be of benefit to subjects, has been criticized as undermining the integrity of the concept [18]. Efforts to reach a consensus definition have been difficult and in many eyes unsatisfactory [3, 19]. Indeed, Goldberg has argued that there will always be difficulty in arriving at necessary and sufficient criteria for TM since it consists of a set of varied manifestations that bear only a "family resemblance" to each other [19]. At best, then, any attempt to define and operationalize the concept can strive to be plausible, but is unlikely to be definitive.

In addition to the definitional uncertainties, research has been plagued by the lack of a validated measure of TM, however defined. As Henderson and colleagues have noted, most studies have been based on open-ended interviews rather than standardized questions, making data gathering and analysis time-consuming and difficult to replicate, and the few studies that have created scales to assess TM have not attempted to validate their measures

[3]. Although in-depth interviews that assess subjects' broader understanding of research participation may be the most definitive means of identifying TM [20], a validated scale to identify subjects who appear to manifest some degree of TM would be a major step forward. The absence of such an instrument not only has hindered assessment of the prevalence and characteristics of the phenomenon, but also the development of meaningful efforts to reduce TM, because of problems in measuring the effectiveness of such interventions. As a consequence, some commentators have concluded, arguably prematurely, that TM is an inevitable and intractable concomitant of consent to clinical research [8],[21].

The primary goal of this study, therefore, was to develop a plausible, theoretically grounded measure of TM, to validate it on a large and diverse sample of research subjects, and to assess its diagnostic accuracy. By doing so, we hoped both to test previous findings regarding the prevalence of TM with a more rigorous methodology and to facilitate future studies aimed at reducing the prevalence of TM by offering a means of assessing the potency of proposed remedies.

Methods

Participants

Two hundred twenty participants were recruited from clinical trials at four academic medical centers in different regions of the U.S., through referrals from principal investigators (PIs) or their research staff. Recruitment occurred from December 2009 to May 2011. Appropriate trials were identified by contacting IRBs, clinical trials offices, and PIs at each site, as well as by searching clinicaltrials.gov. PIs who agreed to participate were asked to query newly enrolled, eligible research subjects about their willingness to be interviewed for this study and to forward contact information for those who agreed. Eligibility criteria included being English-speaking and at least 18 of age, and having signed consent to participate in a randomized intervention trial within the past two months. Subjects were interviewed either in person or by telephone, after their informed consent was obtained for participation. Because potential subjects were referred from the clinical trials in which they were enrolled, we are unable to ascertain the number of subjects whom investigators failed to ask about participation in this study or the number who declined to be contacted. Procedures for the study were approved by the institutional review board at each site.

Assessment of TM: TM questionnaire

Participants were asked to complete a 28-item Likert-type questionnaire to assess the presence of beliefs associated with TM. The theoretical framework for the questionnaire was derived from previous work by two of the investigators that identified two dimensions associated with the phenomenon: unreasonable beliefs, based on a misunderstanding of the methods of the research, in 1) the degree of individualization of the intervention being provided and 2) the likelihood of benefit from participation [2]. A third dimension, misunderstanding of the purpose of research as intended to benefit future patients, was drawn from the effort to develop a consensus definition of TM by Henderson et al. [3]. Although there is not a firm consensus in the literature regarding the definition of TM, all of the alternative definitions of which we are aware draw on one or more of these 3 components.

Each dimension was included because of its potential to compromise the meaningfulness of subjects' consent. Subjects who mistakenly believe that interventions will be individualized for their needs or who hold mistaken beliefs, based on a misunderstanding of the methods of the research, about the likelihood of personal benefit misunderstand how being in research differs from ordinary treatment. Subjects who do not fully understand that the primary

purpose of research is to collect generalizable data to help patients in the future fundamentally misconstrue the nature of the situation into which they are entering and thus base their decisions on incorrect premises [7]. This may be true even if subjects recognize that collecting data is one of several goals of the study, if they simultaneously fail to acknowledge its primacy, since in reality the priority given this goal will determine how decisions about their care in the study are made.

The questionnaire included 3–4 items for each of the three theoretical dimensions (individualization, benefit, and purpose) at three different levels of application (research in general, the project in which the participant was enrolled, and the participant's own treatment), for a total of 28 items. Previous research had demonstrated that some people who recognize the nature of research procedures at one of the more general levels nevertheless fail to demonstrate adequate appreciation of how research procedures affected their own situation, thus reducing the quality of their consent and indicating the presence of TM [1]. Since not all items were applicable to every research study, interviewers were instructed to omit certain items when appropriate (e.g., an item about limitations on adjunctive treatments if no limitations were present in the study).

Assessment of TM: TM Interview

To validate the TM questionnaire, participants also participated in a semi-structured interview designed to elicit their perceptions of the nature of the research in which they were enrolled, the current "gold standard" for the assessment of TM [3]. Questions encouraged participants to discuss their views of the extent to which decisions about their treatment would be based on their individual needs; their expectations of benefit from the study and the reasons for them; and their understanding of the purpose of the study. (See Appendix A) Interviewers were instructed to probe adequately to allow subjects' responses to be scored on these 3 dimensions.

Data collection

Prior to the beginning of data collection, interviewers from all sites participated in an intensive, three-day training session at the coordinating site for the study. The PI and the project director, both of whom had extensive experience with previous studies on TM in clinical research, conducted the training. Procedures included demonstrations of the use of the TM questionnaire and interview, with each interviewer having the opportunity to conduct supervised mock interviews for practice and feedback. During the course of the study, there were regular conference calls involving the interviewers and the project director to attempt to maximize consistency in data collection across sites.

The TM questionnaire was administered verbally. Participants were instructed to do their best to respond to the items without additional verbal clarification, as if they were completing the survey independently. The semi-structured interview followed the questionnaire. Study procedures took approximately 45 minutes to complete.

Data analysis: TM questionnaire

The goals of the statistical analyses were consecutively to establish the factor structure of the TM questionnaire, eliminating those items that failed to achieve adequate factor loadings; to determine the reliability and validity of the questionnaire; to evaluate the diagnostic accuracy of empirically selected cut-offs for the new scale; and to correct for potential optimistic bias in the operating characteristics of the scale through internal cross-validation. Data quality was evaluated through examination of individual item descriptive characteristics. The mean and distribution of scores were examined for each item.

The dimensionality of the items was examined using confirmatory factor analysis (CFA) [22] in Mplus, v.3.11 [23], using the following models: a one-factor model, representing TM; a 3-factor model with correlated dimensions; and a hierarchical 3-factor model. We excluded items with poor factor loadings (<.600) and compared the fit of the 3 factor structures using a chi-square difference test. In line with recommendations, model fit was evaluated using several fit indices for each model [24], including the chi-square statistic, the Comparative Fit Index (CFI), the root mean squared error of approximation (RMSEA), and the Tucker-Lewis Index (TLI). The CFI and the TLI are incremental fit indices that measure model fit by comparing the specified model with a baseline model where the observed variables are mutually uncorrelated. Hu and Bentler [24] suggest CFIs should be greater than .95 for good model fit and greater than .90 for acceptable fit, and a cutoff value of .95 for the TLI. The RMSEA is a fit index based on the function of error of approximation of the best fitting model to the population covariance matrix [25]. The widely used RMSEA cut-off scores are zero for perfect fit, .05 for good fit and .08–.10 or less for acceptable fit, but recent work has suggested that these cutoffs are rather arbitrary and are influenced by sample size and model specifications [26].

The internal consistency of the TM scale retained in the final CFA model was evaluated by Cronbach's alpha, with .70 indicating acceptable reliability [27]. To assess the external validity of the TM scale, the means of subjects coded as having or not having TM based on the TM interview were examined. Subjects classified as having TM by the interview were expected to have significantly higher scores.

We used a logistic regression model with the TM scale score as a predictor and the coded "gold standard" of the TM interview as an outcome to examine the, receiver operating curve (ROC) of the scale and to determine the optimal cut points on the TM scale. A cut point is usually the score that best strikes a balance between the sensitivity (proportion of true positives) and specificity (proportion of true negatives) of a scale, although depending on the use of the scale different cut points may be selected, for example to maximize the sensitivity of a tool [28]. For the ROC plots the sensitivity and specificity values for each score on the TM scale were estimated and the values were plotted, representing graphically the performance of the scale through the range of cut points. The area under the ROC curve (AUC) was calculated to summarize the screening/ diagnostic utility of the TM scale. At the level of chance, the AUC will be 0.5 and is represented by the diagonal on the ROC graph. The AUC increases as the accuracy of the test increases to a maximum of 1.0, which indicates a test with perfect diagnostic utility. Tests with AUC in the .7–.9 range are considered moderately accurate [28].

Ideally the selection of a cut-off point is informed by the prevalence in the target population and the relative consequences of false positive and false negative test results, which can vary depending on the context [28]. As this study is an initial evaluation of a proposed scale we took an empirical approach, assuming 50% prevalence of the condition and equivalent costs of false-positive and false-negative results and evaluating the diagnostic accuracy of a cut-off based on the Youden Index [29], which optimizes the sensitivity and the specificity of the scale. In addition we examined cut-off points corresponding to 80% sensitivity and to 80% specificity.

Given that diagnostic accuracy of tests and cut-off values based on empirical investigation are prone to an optimistic bias, we conducted a 10-fold internal cross-validation to acquire more realistic estimates of test performance. We randomly divided the data into tenths, iteratively trained the model on 9/10 of the data, and calculated the predicted probabilities of presence of TM (interview results) for each observation based on the derived model in the remaining 1/10 test set. We used these results to plot an ROC and compute the areas under

the curve to evaluate the global diagnostic accuracy of the measure [30–32]. In addition we evaluated the diagnostic accuracy of cut-off scores selected on the basis of the Youden Index, and at 80% sensitivity and 80% specificity. To evaluate the performance of the test at these cut-offs, we examined sensitivity (the ability of the test to detect the condition when it is truly present) and specificity (the ability of the test to exclude the condition in patients who do not have it), positive predictive value (probability that a person has TM, when the tests are positive; PPV), negative predictive value (probability that a person does not have TM when test results are negative; NPV); the likelihood ratio for a positive result (how much the odds of the condition increase with a positive test; +LR); and the likelihood ratio for a negative result (how much the odds of the condition decrease with a negative test; -LR). The training data sets were also used to evaluate the stability of the derived factor structure and internal validity of the TM scale.

Data analysis: TM interview

All interviews were transcribed and coded for TM. The coding system was based on a previous study by two of the authors [2]. When comparing codes across coders, wherever agreement was less than 100%, code definitions were reviewed, clarified and refined as needed to increase coding reliability. The coder, a bachelor's-level research coordinator with previous qualitative data coding experience, and the Project Director, an experienced master's-level member of the research team, then independently coded sets of 6 to 8 transcripts at a time. This was followed by meetings with the PI, who has three decades of experience coding transcripts for TM, where each transcript was reviewed and any differences in coding discussed until a consensus was reached.

For each case, coders reviewed the transcripts to assess the 3 dimensions of TM described above. Coders were permitted to use information from answers to the open-ended questions only; no responses derived from the verbally administered TM questionnaire were used and coders were blind to the results of the questionnaires.

Evidence of TM on each dimension was ascertained using the following rules and coded as present, absent, or insufficient evidence:

Individualization—Any clear evidence of a belief that treatment choices would be individualized for the subject's specific needs, when that belief was inconsistent with the study protocol.

Example (subject 414 - Phase 3 randomized trial of chemotherapy for metastatic adenocarcinoma of the pancreas):

Interviewer: And you talked a little bit about randomization; about how they decide. Any more details about that?

Participant: No. All I know is they take your studies [lab results] and they put them into a computer, I guess. And they put all the factors in the computer and the computer comes up with, okay, he should be on this. They put the information in there and it comes up with, through prior research, it'll tell them what you should be getting…according to the data they've got on you.

Interviewer: In labs and things.

Participant: Right.

Benefit—Any clear evidence that both of the following were present: 1) a belief that there was likely to be personal therapeutic benefit from participation *and* 2) the methods of the study precluded the perceived benefit from occurring or the efficacy of the experimental medication or intervention was unproven. In studies where everyone received standard treatment and subjects were then randomized to an experimental add-on treatment or placebo, TM was scored as present only if the subject expressed a strong conviction that the benefit was likely to accrue as a result of the experimental (i.e., unproven) component and/or a certainty that they would get the experimental treatment. Expression of a subject's hope for benefit was not scored as TM, in contrast to cases in which a subject characterized such benefit as likely.

Example: (subject 118- Phase 3 randomized, double-blind, placebo-controlled, multicenter study of medication for digital lesions in scleroderma)

I: Uh huh. And how likely do you think it is that you are gonna benefit from being in the study?

P: I feel pretty sure that it's gonna help me and mainly right now my goal is to get my fingers better. I can't do anything without my hands. I've got to have them.

Purpose of the study—Any clear evidence of a belief that the *primary* purpose of the study was to help the study participants, rather than to help patients in the future or to attain another scientific goal.

Example (subject 318 - Phase 3 trial for recurrent or metastatic breast cancer):

Interviewer: And would you say the study is primarily designed to help the participants in the study or to collect data to help people in the future?

Participant: The way I have…my experience is to treat the patient. It is mentioned to you that that it is a study and that the information will be used for the betterment of others. Yeah, but in this case, the focus is on the patient.

In addition to coding the interviews for the evidence of TM on each dimension, a composite category was created indicating evidence of TM on any dimension. In keeping with our usual procedure, any clear evidence of TM was coded as TM present, even if at some other point in the interview the subject gave a contradictory response; we adopt this approach due to inherent uncertainty regarding the impact of subjects' contradictory views on their decision making, our desire to be inclusive with regard to cases in which TM may negatively affect the quality of consent, and our experience that TM beliefs are often central to decisions about participating in a trial [1]. Cases where interviewers coded no TM for some dimensions and could not agree on its presence or absence on others (33 cases for the purpose dimension, 19 for benefit, 36 for individualization, 2 for all dimensions) were coded as having no evidence of TM. Only cases where coders could not reach agreement on any of the dimensions were coded as having insufficient evidence to make a decision about the presence of TM. These cases were excluded from the validity and ROC analyses, leaving 189 participants with available TM score and interview data for these analyses.

Results

The characteristics of participants are shown in Table 1. They were predominantly white and male, with at least some college education, actively working or retired, and spread across the age spectrum. Table 2 shows the nature of the studies in which participants were enrolled. Psychiatry (28%; n=62) and hematology/oncology (28%; n=61) made the biggest

contributions, followed by neurology (12%; n=27); the remaining subjects were drawn from a wide range of areas of medicine. Most studies (59.5%; n=50) were placebo-controlled and most subjects (68%; n=150) came from these trials; and most trials were Phase 2 or Phase 3 (75%; n=63).

Endorsement of items associated with TM was common among participants in the study, as shown in Table 3, but varied across items. On the TM interview, 24.50% (n=49) of participants were coded as showing evidence of TM on the individualization dimension, 35% (n=70) on the benefit dimension, and 15% (n=30) on the purpose dimension. 50.5% (n=101) of participants presented evidence of TM on at least one dimension.

Factor Analysis of TM Questionnaire

All items on the questionnaire were recoded so that higher scores would indicate stronger responses consistent with TM. Examination of descriptive characteristics revealed no items with out of range values. Two of the items evaluating purpose (Table 3; items P3 and P5) were found to be highly skewed and were excluded from further analyses. Five of the items (Table 3; items I1, I2, I4-I7) were not administered to all participants, as they were deemed not to be relevant to specific study designs. Since inclusion of these items would present challenges in interpretation and application by future users of the scale, they were omitted from these analyses. All remaining items were included in the initial CFA model.

Based on the preselected cut-off, items with factor loadings <.6 were excluded from the analyses in a step-wise fashion (see Table 3 for excluded items). Ten additional items were excluded based on this criterion and the final CFA models were evaluated on the 10 items with good factor loadings. The one factor model demonstrated acceptable fit (χ^2 (21) = 71.31, p<.0001, CFI = .961, TLI =.983 and marginal RMSEA=.11). The 3-factor model had slightly better fit indices (χ^2 (19) = 55.5, p<.0001, CFI = .972, TLI =.987 and RMSEA=.10) and very high correlations among factors (range .887–.969). The chi-square difference test comparing the two models was significant (χ^2 (3) = 20.8, p<.0001) suggesting that the 3-factor model provides a better fit for the data (Table 4). The final model supported the hypothesized presence of 3 related TM domains, which were strongly correlated. The hierarchical model failed to converge in this sample. The resulting TM scale had an excellent internal consistency (Cronbach's alpha=.90).

The fit indices of the model replications across the 10 training sets were consistent with the original results (CFI range .963–.974, RMSEA range .987–999), suggesting a stable factor solution.

External Validity and ROC Analyses of the TM Scale

As the 3 domains in the final CFA model were highly correlated, a single TM scale score was computed by summing the 10 items retained in the model. To evaluate the validity of the scale, we tested its ability to differentiate between participants who were identified as having evidence of TM on the TM interview from those who were not. The TM scores were significantly higher for participants who had interview results indicating presence of TM on each dimension and on all dimensions combined (Table 5), confirming the external validity of the scale. The subscale scores of the TM scale (benefit, individualization, purpose) were also significantly higher for participants with interview results indicating presence of TM on the corresponding dimension (data not shown).

The results of the ROC plots indicated that, based on the Youden Index, 27 is the cut-off point on the TM scale that would maximize the sensitivity and the specificity of the measure; with that cut-off, 55% of the sample would be categorized as manifesting TM. A score of 24 would be the cut-off value for 80% sensitivity (proportion of sample with TM

would be 62%) and a score of 35 would be the cut-off value for 80% specificity (35% with TM). The diagnostic accuracy measures for these points based on the original model and the 10-fold internal cross-validation results are presented in Table 6. The cut-off value selected may depend on the intended use of the scale in a given instance and the relative cost of false positive and false negative classifications. The ROC curve in Figure 1 represents the performance of the TM scale across all cut-off points for the combined category indicating any evidence of TM in the interview. The value of the AUC, used to summarize the diagnostic utility of a test was .703 (95% CI: .627 – .778), suggesting moderate accuracy of the scale [28]. Results from the 10-fold internal cross-validation corrected for some optimistic bias, leading to an adjusted AUC of .682 (95% CI: .605 – .706). Although the two ROCs differed significantly, the change in magnitude of the AUC was only marginal. Some small magnitude correction was also evident in the diagnostic accuracy measures of the scale at different cut-off points (Table 6).

Overall the TM scale demonstrated modest predictive value and likelihood ratios. The LR is important measure of diagnostic accuracy. Unlike the PPV and NPV, it does not depend on the prevalence of a condition and it takes into account all available information in a classification table. Positive and negative LRs are also very helpful in clinical practice, as they can be used to predict the posttest probability of a condition for a given patient when the pretest probability is known. For example, based on the +LR of 1.89 for the TM cutoff of 27 and a population with a prevalence of TM of 60%, the probability of TM in a patient with TM score >27 would increase to 74% (from 60%). For a patient with a TM score of <27 the probability of TM would decrease to 41%, based on the –LR of .47.

Discussion

The final 10-item TM scale derived in this study of 220 participants in a wide range of clinical trials is based on data suggesting that TM can be manifest by mistaken beliefs on one or more of 3 dimensions: the degree of individualization of treatment, likely benefit from a study, and the overall purpose of the study. Elimination of problematic items and of items that failed to achieve adequate factor loadings from the original 28 items in the questionnaire resulted in a 10-item scale that loaded on to 3 highly correlated factors, corresponding to each of the original 3 dimensions. The scale had excellent internal consistency and showed good ability to discriminate between participants who displayed manifestations of TM on the TM interview and those who did not. The proposed TM scale demonstrated diagnostic accuracy in the lower moderate range. The AUC for the TM scale was lower than the AUCs reported for some well-established self-report measures used as diagnostic screeners like the PHQ-9 [33] and BDI [34], where a relatively robust gold-standard criterion is available for comparison. On the other hand, the TM scale's overall accuracy was within the range of global accuracy reported by other self-report scales of constructs assessing states of mind rather than discrete symptoms [35–36] or by established scales in new patient groups [37]. In addition most studies report measures of diagnostic accuracy without correction for possible optimistic bias. Results from the internal validation of the TM scale we conducted suggested modest correction for optimistic bias, which is consistent with results reported from simulation studies with similar sample size [38] and suggests that further reduction in diagnostic accuracy measures is unlikely.

Data from both the TM interview and the TM scale confirm previous suggestions that evidence of TM is commonly found among research subjects. The frequency with which participants manifested some evidence of TM was 50.5% based on the TM interview and 55% based on the TM scale using a cutoff score of 27. Both analyses suggest somewhat lower rates of TM than prior studies, which found evidence for TM in 60–74% of research subjects [2],[9]-[12], using a wide variety of definitions and methods of assessment. The

variation in rates may be due to differences in samples, definitions of TM, assessment approaches, or the effectiveness of researchers' informational presentations to patients. Although the nature of subject referral to the study prevents us from knowing how many subjects who were approached by the research staff of their primary studies declined to be interviewed, the rates of TM in this study suggest that investigators were not systematically excluding their less well-performing subjects.

Our findings may help to explain some of the difficulty that the field has had in converging on a common definition of TM, while at the same time generating remarkably similar estimates of its prevalence across diverse samples. These data support the notion that TM has diverse manifestations along at least 3 dimensions. Different research groups appear to have focused on one or more of those dimensions, often to the exclusion of the other(s). However, since the dimensions are strongly correlated, the resulting estimates of the prevalence of TM showed great similarity. As Goldberg suggested, TM appears to manifest itself as a set of phenomena that display a familial resemblance, [19] akin to the concept of a "fuzzy set." [39] It seems likely, though, that the most accurate assessments of TM will take into account all three of the dimensions identified in this study as contributing to the concept. This should encourage convergence on a uniform approach to the conceptualization of TM.

Inherent in the construction of a scale to assess TM is the belief—confirmed by our data—that the phenomenon exists along a spectrum of intensity. A small number of subjects manifest a great deal of TM, permeating their appreciation of every element of the study. Other subjects show only focal deficits related to one or another element. Many subjects lie somewhere in between. Where to draw the line at which someone has "enough" TM to be excluded from offering consent is a matter of policy, and may vary according to the risk-benefit profile presented by a study. Exactly the same is true for competence to consent to research participation or treatment; the most commonly used measures of competence provide quantitative estimates of the degree of impairment on the basis of which a categorical choice is made regarding the acceptability of the subject's consent, usually based on the risks and benefits of research participation or treatment [40]. In both cases, as risk increases, a higher level of certainty that the subject is offering a meaningful consent is desirable.

It is important to keep in mind precisely what a scale of this sort can and cannot do. By providing evidence of beliefs associated with TM, the scale can help to identify research participants who—to some degree—fail to appreciate the nature of the study into which they are being asked to enter and the probable consequences of their involvement. Moreover, by offering a measure of the number of mistaken beliefs, the scale may allow investigators to target for additional education those subjects at highest risk for distorted decision making. However, as Kim and colleagues have noted, subjects' responses to a standard set of questions about their beliefs may be misleading for at least 3 reasons: subjects may not understand the questions themselves; investigators may misinterpret subjects' responses; and even when subjects display some misconceptions about studies, the confusion may not actually impact their decisions. [20] They suggest that only in-depth interviews can overcome these problems, although it is clear that such an approach is not a panacea either, since subjects' responses may be inherently ambiguous, confounding efforts at coding. Though in-depth interviews represent the current "gold standard," the time required to conduct and code such interviews renders them impractical for use in clinical trials. Hence, this scale can best be understood as a more efficient mechanism for ascertaining which subjects are at risk for distortions in their decisions and to what degree, but will not allow researchers to conclude definitively that TM is present in a given subject or that, if present, it will necessarily impact the subject's decision. Some followup with identified subjects will

be necessary, and the use of the scale should not replace individualized inquiry when subjects' comprehension is in doubt. Moreover, a replication of our findings regarding the reliability and validity of the scale, especially when used by other research groups, would be appropriate.

With these cautions taken into account, the TM scale described here may be useful for a variety of purposes. *Clinical researchers* who would like to identify participants who may harbor serious misconceptions about the nature and consequences of the study to which they are being asked to consent may find it helpful to use the TM scale to screen for such participants, who can then receive additional education; for this purpose, researchers may select a cutoff that maximizes the sensitivity of the scale, even at the cost of reduced specificity, prioritizing the identification of those subjects who would benefit from supplementary information about clinical research. *IRBs and other research ethics committees*, concerned that subjects in higher risk clinical studies might not appreciate the implications of research participation, might ask researchers to use the TM scale to establish eligibility for enrollment or for referral for further education prior to enrollment; here, a cutoff that balances specificity and sensitivity may be more appropriate, given the importance of minimizing the number of subjects wrongly identified as manifesting TM and hence ultimately excluded from participating. It may also be helpful in educating investigators and research staff about the nature and prevalence of TM, and its potential impact on informed consent. Investigators who study the process of informed consent to clinical trials and other clinical studies may find this scale of use in ascertaining evidence of TM and, even more importantly, in assessing the effectiveness of efforts to reduce its manifestations. Indeed, given a good deal of data accumulated over several decades suggesting a high prevalence of TM in clinical research, we believe that future studies should focus on what can be done to correct subjects' misconceptions.

Acknowledgments

References

1. Appelbaum PS, Roth LH, Lidz CW, Benson P, Winslade W. False hopes and best data: Consent to research and the therapeutic misconception. Hastings Cent Rep. 1987; 17(2):20–24. [PubMed: 3294743]

2. Appelbaum PS, Lidz CW, Grisso T. Therapeutic misconception in clinical research: frequency and risk factors. IRB Ethics Hum Res. 2004; 26(2):1–8. correction and clarification (2004) 26(5):18.

3. Henderson GE, Churchill CR, Davis AM, Easter MM, Grady C, et al. Clinical trials and medical care: defining the therapeutic misconception. PLoS Med. 2007; 4:1735–1738.

4. Appelbaum, PS.; Lidz, C. The therapeutic misconception. In: Emanuel, EJ.; Grady, C.; Crouch, RA.; Lie, RK.; Miller, FG.; Wendler, D., editors. The Oxford textbook of clinical research ethics. New York: Oxford University Press; 2008. p. 633-644.

5. Henderson GE, Easter MM, Zimmer C, King NMP, Davis AM, et al. Therapeutic misconception in early phase gene transfer trials. Soc Sci Med. 2006; 62:239–253. [PubMed: 16000230]

6. Horng S, Grady C. Misunderstanding in clinical research: distinguishing therapeutic misconception, therapeutic misestimation, and therapeutic optimism. IRB Ethics Hum Res. 2003; 25(1):11–16.

7. Miller FG, Joffe S. Evaluating the therapeutic misconception. Kennedy Inst Ethics J. 2006; 16:353–366. [PubMed: 17847601]

8. Sreenivasan G. Does informed consent to research require comprehension? Lancet. 2003; 362:2016–2018. [PubMed: 14683665]

9. Kimmelman J. The therapeutic misconception at 25: treatment, research and confusion. Hastings Cent Rep. 2007; 37(6):36–42. [PubMed: 18179103]

10. Dunn LB, Palmer BW, Keehan M, Jeste DV, Appelbaum PS. Assessment of therapeutic misconception in older schizophrenia patients using a brief instrument. Am J Psychiatry. 2006; 163:500–506. [PubMed: 16513873]

11. Wazaify M, Khalil SS, Silverman HJ. Expression of therapeutic misconception amongst Egyptians: a qualitative pilot study. BMC Med Ethics. 2009; 10(7) Available: http://www.biomedcentral.com/1472-6939/10/7.

12. Durand-Zaleski IS, Alberti C, Durieux P, Duval X, Bottot S, et al. Informed consent in clinical research in France: assessment and factors associated with therapeutic misconception. J Med Ethics. 2008; 34(9):e16. Available: http://jme.bmj.com/content/34/9/e16.full.pdf. [PubMed: 18757616]

13. Joffe S, Cook EF, Clearly PD, Clark JW, Weeks J. Quality of informed consent in cancer clinical trials: A cross-sectional survey. Lancet. 2001; 358:1772–1777. [PubMed: 11734235]

14. Snowdon C, Garcia J, Elbourne D. Making sense of randomization: Responses of parents of critically ill babies to random allocation of treatment in a clinical trial. Soc Sci Med. 1997; 45:1337–1355. [PubMed: 9351153]

15. Advisory Committee on Human Radiation Experiments. Final report of the advisory committee on human radiation experiments. New York: Oxford University Press; 1996.

16. Hereu P, Perez E, Fuentes I, Vidal X, Sune P, et al. Consent in clinical trials: what do patients know? Contemp Clin Trials. 2010; 31:443–446. [PubMed: 20462521]

17. Jansen L, Appelbaum PS, Klein W, Sulmasy D, Weinstein N, et al. Unrealistic optimism in early phase oncology trials. IRB Ethics Hum Res. 2011; 33(1):1–8.

18. Kimmelman J. The therapeutic misconception at 25: treatment, research, and confusion. Hastings Cent Rep. 2007; 37(6):36–42. [PubMed: 18179103]

19. Goldberg DS. Eschewing definitions of the therapeutic misconception: a family resemblance analysis. J Med Philos. 2011; 36:296–320. [PubMed: 21606116]

20. Kim SYH, Schrock L, Wilson RM, Frank SA, Holloway RG, et al. An approach to evaluating the therapeutic misconception. IRB Ethics Hum Res. 2009; 31(5):7–14.

21. Dawson, A. What should we do about it? Implications of the empirical evidence in relation to comprehension and acceptability of randomization. In: Holm, S.; Jonas, M., editors. Engaging the world: the use of empirical research in bioethics and the regulation of biotechnology. Amsterdam: IOS Press; 2004.

22. Embretson, SE.; Reise, SP. Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum Associates, Inc; 2000.

23. Muthen, LK.; Muthen, BO. Mplus user's guide. Los Angeles, CA: Muthen & Muthen; 1998–2004.

24. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999; 6(1):1–55.

25. Steiger, JH.; Lind, J. Statistically-based tests for the number of common factors; Paper presented at the annual spring meeting of the Psychometric Society; Iowa City. 1980.

26. Chen F, Curran PJ, Bollen KA, Kirby J, Paxton P. An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. Sociol Methods Res. 2008; 36(4):462–494. [PubMed: 19756246]

27. Bland JM, Altman DG. Statistical notes: Cronbach's alpha. BMJ. 1997; 314:572. [PubMed: 9055718]

28. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med. 2000; 45(1–2):23–41. [PubMed: 10802332]

29. Youden WJ. Index for rating diagnostic tests. Cancer. 1950; 3(1):32–35. [PubMed: 15405679]

30. Altman DG, Bland JM. Diagnostic tests 1: sensitivity and specificity. BMJ. 1994; 308(6943):1552. [PubMed: 8019315]

31. Altman DG, Bland JM. Diagnostic tests 2: predictive values. BMJ. 1994; 309(6947):102. [PubMed: 8038641]

32. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ. 1994; 309(6948):188. [PubMed: 8044101]

33. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001; 16(9):606–613. [PubMed: 11556941]

34. Cameron IM, Cardy A, Crawford JR, du Toit SW, Hay S, Lawton K, Mitchell K, Sharma S, Shivaprasad S, Winning S, Reid IC. Measuring depression severity in general practice: discriminatory performance of the PHQ-9, HADS-D, and BDI-II. Br J Gen Pract. 2011; 61(588):e419–e426. [PubMed: 21722450]

35. Asadi-Lari M, Packham C, Gray D. Is quality of life measurement likely to be a proxy for health needs assessment in patients with coronary artery disease? Health Qual Life Outcomes. 2003; 1(1):50. [PubMed: 14596682]

36. Smith S, Trinder J. Detecting insomnia: comparison of four self-report measures of sleep in a young adult population. J Sleep Res. 2001; 10(3):229–235. [PubMed: 11696076]

37. Hedayati SS, Bosworth HB, Kuchibhatla M, Kimmel PL, Szczech LA. The predictive value of self-report scales compared with physician diagnosis of depression in hemodialysis patients. Kidney Int. 2006; 69(9):1662–1668. [PubMed: 16598203]

38. Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. Clin Chem. 2008; 54(4):729–737. [PubMed: 18258670]

39. McNeill, D.; Freiberger, P. Fuzzy logic: the revolutionary computer technology that is changing our world. New York: Simon and Schuster; 1994.

40. Dunn LB, Nowrangi MA, Palmer BW, Jeste DV, Saks ER. Assessing decisional capacity for clinical research or treatment: a review of instruments. Am J Psychiatry. 2006; 163(8):1323–1334. [PubMed: 16877642]

Appendix A - TM Interview Questions

Purpose

What is your understanding of the purpose of the study?

Suggested probes (if needed):

Why are the doctors doing this study?

Is the study primarily designed to help participants in the study or to collect data to help people in the future?

Benefit

How do you think that being in the study might (might have) help(ed) you?

Suggested probes (if needed):

How likely do you think it is that you'll benefit from being in this study?

In what ways? What makes you think that?

Are there any disadvantages to being in the study?

Individualization

How would (will) your personal treatment be different if you were (since you are) not in this study?

Suggested probes (if needed):

How will decisions about your treatment be made in this study?

How will it be decided who gets what treatment?

Are there any restrictions on the treatment the research doctors can give you as a result of your being in the study?
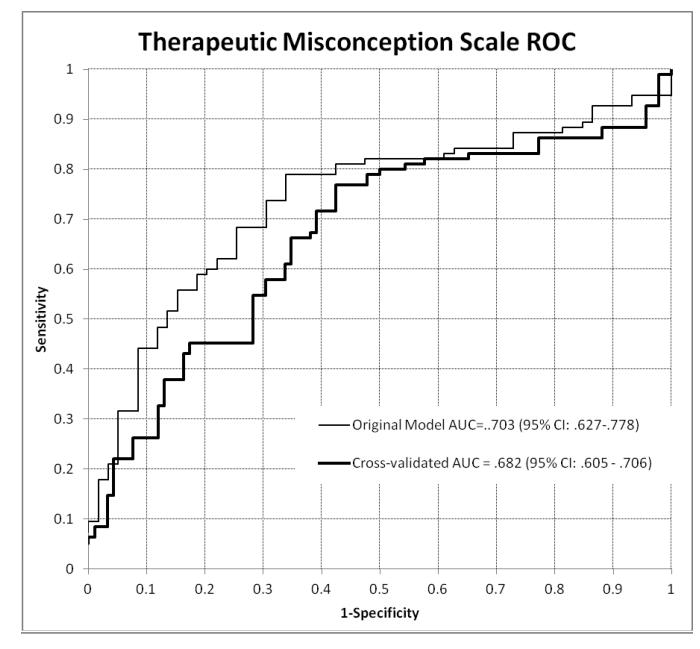
Figure 1.

Table 1

Participant characteristics

		All recruited (N=220)		Included in validity analyses (N=189)	
		n	%	n	%
Age					
	<30	38	17.27	35	18.52
	30–39	25	11.36	22	11.64
	40–49	32	14.55	26	13.76
	50–59	52	23.64	45	23.81
	60–69	52	23.64	41	21.69
	70+	20	9.09	19	10.05
Female		97	44.09	85	44.97
Hispanic		22	10	18	9.52
Race					
	American Indian or Alaska Native	2	0.91	1	0.53
	Asian	8	3.64	8	4.23
	Black or African American	25	11.36	22	11.64
	White	166	75.45	146	77.25
Employment status					
	Currently employed	115	52.27	103	54.50
	Retired	46	40.35	41	40.20
	Looking for work	14	12.28	14	13.73
	Keeping house/raising children	7	6.14	5	4.90
	Other	46	40.35	41	40.20
Education					
	Some High School	9	4.11	9	4.79
	High School Diploma	26	11.87	19	10.11
	Completed GED	6	2.74	6	3.19
	Voc/Trade/Business School/Military Training	15	6.85	14	7.45
	Some College or a 2-Year Degree	62	28.31	54	28.72
	Finished a 4-Year Degree	59	26.94	51	27.13

	All recruited (N=220)		Included in validity analyses (N=189)	
	n	%	n	%
Master's Degree or Equivalent	24	10.96	22	11.70
Other Advanced Degree	17	7.76	12	6.38

Table 2

Clinical Trials from Which Participants Were Recruited

	Trials (N=84)		All participants recruited (N=220)		Included in validity analysis (N=189)	
	n	%	n	%	N	%
Field of Medicine						
Cardiovascular Medicine	1	1.2	2	0.9	2	1.1
Dentistry/Periodontics	1	1.2	4	1.8	1	0.5
Dermatology	1	1.2	4	1.8	3	1.6
Endocrinology	5	5.9	6	2.7	5	2.6
Family/Complimentary Medicine	1	1.2	2	0.9	2	1.1
Gastroenterology	2	2.4	2	0.9	1	0.5
General Internal Medicine	2	2.4	6	2.7	5	2.6
Hematology/Oncology	30	35.7	61	27.7	50	26.5
Infectious Diseases and Immunology	2	2.4	5	2.3	5	2.6
Nephrology	1	1.2	1	0.5	0	0
Neurology	13	15.4	27	12.3	22	11.6
Obstetrics and Gynecology	1	1.2	3	1.4	3	1.6
Ophthalmology	1	1.2	1	0.5	1	0.5
Orthopedics	1	1.2	3	1.4	3	1.6
Psychiatry	15	17.8	62	28.1	57	30.2
Pulmonary Medicine	3	3.6	11	5	10	5.3
Rheumatology	3	3.6	18	8.2	17	9
Surgery	1	1.2	2	0.9	2	1.1
Trial Phase						
I & II	5	5.9	20	9.1	19	10
II	35	41.7	87	39.5	74	39.2
II & III	1	1.2	2	0.9	2	1.1
III	37	44	78	35.5	66	34.9
IV	5	5.9	32	14.5	28	14.8

	Trials (N=84)		All participants recruited (N=220)		Included in validity analysis (N=189)	
Unknown	1	1.2	1	0.5	0	0
*Placebo Controlled Trials	50	59.5	141	64	127	67.2

*
Of the 50 trials with placebos, 27 were "add-on" studies, in which standard treatment was provided to all participants, in addition to which some received an experimental intervention and others placebo.

Table 3

Item descriptive characteristics

Item text	Label		1	2	3	4	5	Mean	SD	Exclusion reason[#]
			Frequencies by response category							
People in this study may not do as well as they would in usual treatment.	B1	n	28	26	17	40	103	3.77	1.48	C
		%	13.08	12.15	7.94	18.69	48.13			
People in this study will do better than they would if they were just getting treatment as usual. ^	B2	n	55	37	21	51	49	3.01	1.54	C
		%	25.82	17.37	9.86	23.94	23			
Ordinary treatment could turn out to be better than the treatment people receive in this study.	B3	n	66	37	12	31	67	2.98	1.68	C
		%	30.99	17.37	5.63	14.55	31.46			
My own treatment for [disorder] will almost certainly be better as a result of participating in this study. ^	B4	n	51	19	21	46	79	3.38	1.60	C
		%	23.61	8.8	9.72	21.3	36.57			
The reason I was asked to be in this study is that it will provide me with the best treatment available. ^	B5	n	81	28	9	38	61	2.86	1.71	
		%	37.33	12.9	4.15	17.51	28.11			
The treatment I am getting by being in this study is the best treatment for me. ^	B6	n	54	25	27	34	76	3.25	1.62	
		%	25	11.57	12.5	15.74	35.19			
There are other treatments I could get outside this study that might be just as good for me.	B7	n	118	20	22	24	34	2.25	1.56	C
		%	54.13	9.17	10.09	11.01	15.6			
The experimental intervention that people receive in research studies may not be any more effective than those interventions that are not experimental.	B8	n	148	33	8	11	18	1.71	1.26	C
		%	67.89	15.14	3.67	5.05	8.26			
Being in a research study almost always provides the best possible treatment for a sick person. ^	B9	n	91	42	10	40	36	2.49	1.56	
		%	41.55	19.18	4.57	18.26	16.44			
By participating in a research study, people will get the best treatment for their medical problems. ^	B10	n	84	37	15	37	43	2.62	1.60	
		%	38.89	17.13	6.94	17.13	19.91			
People in research studies get treatments that sometimes will not be particularly helpful for the problems that they have.	B11	n	137	38	6	12	23	1.82	1.35	C
		%	63.43	17.59	2.78	5.56	10.65			
This study is being done because the researchers aren't sure whether [medication/treatment/procedure 1*] or	P1	n	117	49	4	17	32	2.08	1.48	C
[medication/treatment/procedure 2*]; add others if	P1	%	54.17	22.69	1.83	7.76	14.81	2.08	1.48	C

necessary] is the best way to treat [disorder].

Item text	Label		Frequencies by response category					Mean	SD	Exclusion reason[#]
			1	2	3	4	5			
This study is designed primarily to help the people in it; more so than to generate knowledge to help future patients with the same condition.[^]	P2	n	126	27	1	26	39	2.20	1.62	C
		%	57.53	12.33	0.46	11.87	17.81			
Although the study may help the participants, researchers are doing this study mainly to figure out which treatment works the best for people with the same condition.	P3	n	164	43	3	5	3	1.35	0.75	A
		%	75.23	19.72	1.38	2.29	1.38			
The purpose of the [experimental study] is to provide the best treatment available for me and the others in the study.[^]	P4	n	58	19	2	32	105	3.50	1.74	
		%	26.85	8.8	0.93	14.81	48.61			
The main purpose of this study is to help people in the future, whether or not it helps me.	P5	n	168	34	3	5	10	1.43	0.98	A
		%	76.36	15.45	1.36	2.27	4.55			
Research studies are designed primarily to help those people who participate in them rather than primarily to help future patients with the same disease[^]	P6	n	148	37	4	13	18	1.71	1.26	
		%	67.27	16.82	1.82	5.91	8.18			
Research is done to gain knowledge for future use. It may help the people who participate but that is not the main goal of the research.	P7	n	152	39	2	8	18	1.63	1.21	C
		%	69.41	17.81	0.91	3.65	8.22			
A researcher's most important task is to make sure that the research will help the people who participate.[^]	P8	n	83	44	1	37	52	2.68	1.67	
		%	38.25	20.28	0.46	17.05	23.96			
This study the amount of medication will be adjusted according to my personal needs.[^]	I1	n	67	9	10	20	62	3.01	1.81	B
		%	39.88	5.36	5.95	11.9	36.9			
Aside from the medication(s)/treatment(s) being studied, I will receive whatever other medication(s)/treatment(s) will be helpful to me.	I2	n	52	6	6	17	66	3.27	1.82	B
		%	35.37	4.08	4.08	11.56	44.9			
This study is designed to give everyone the type of treatment and amount of treatment that best fits his or her individual needs.[^]	I3	n	60	23	6	41	90	3.35	1.70	
		%	27.27	10.45	2.73	18.64	40.91			
I will definitely receive [the experimental medication] in this study.[^]	I4	n	79	4	2	15	47	2.64	1.86	B
		%	53.74	2.72	1.36	10.2	31.97			

Item text	Label		Frequencies by response category					Mean	SD	Exclusion reason[#]
			1	2	3	4	5			
In this study, my doctor will not know exactly what intervention (i.e., experimental medication or another medication) I will be receiving.	15	n	100	14	5	2	20	1.78	1.43	B
		%	70.92	9.93	3.55	1.42	14.18			
The doctors who are providing my care know all the specifics of what [type of intervention] I am getting in the study. ^	16	n	58	10	1	20	54	3.01	1.83	B
		%	40.56	6.99	0.7	13.99	37.76			
The research physician cannot add any other medication/treatment for [disorder] during the [experimental period] of the study, even if he or she thinks it would help me.	17	n	84	17	9	9	25	2.13	1.57	B
		%	58.33	11.81	6.25	6.25	17.36			
When designing research, researchers must be sure that each person in the study will get the best treatment available for that person's individual needs, just as though they were being treated by their personal doctor. ^	18	n	67	29	6	29	88	3.19	1.75	
		%	30.59	13.24	2.74	13.24	40.18			
Researchers always try to provide each person in a study the treatment that best meets that person's individual needs ^	19	n	87	28	4	38	60	2.80	1.73	
		%	40.09	12.9	1.84	17.51	27.65			

Response categories: 1=Agree, 2=Mostly Agree, 3=Don't know, 4=Mostly disagree, 5=Disagree.

^
Items were reverse coded, so that higher score indicates presence of TM.

[#]A- Skewed item, B - Missing responses, C - Poor item loadings

Table 4

Therapeutic misconception scale final confirmatory factor analysis

Item text	Item number	Factor loadings			
		One factor	Three factors		
			B	I	P
1. This study is designed to give everyone the type of treatment and amount of treatment that best fits his or her individual needs.	I3	0.767		0.796	
2. When designing research, researchers must be sure that each person in the study will get the best treatment available for that person's individual needs, just as though they were being treated by their personal doctor.	I8	0.814		0.841	
3. Researchers always try to provide each person in a study the treatment that best meets that person's individual needs	I9	0.855		0.891	
4. The purpose of the [experimental study] is to provide the best treatment available for me and the others in the study.	P4	0.82			0.826
5. Research studies are designed primarily to help those people who participate in them rather than primarily to help future patients with the same disease	P6	0.605			0.609
6. A researcher's most important task is to make sure that the research will help the people who participate.	P8	0.829			0.835
7. The reason I was asked to be in this study is that it will provide me with the best treatment available.	B5	0.794	0.81		
8. The treatment I am getting by being in this study is the best treatment for me.	B6	0.777	0.794		
9. Being in a research study almost always provides the best possible treatment for a sick person.	B9	0.726	0.749		
10. By participating in a research study, people will get the best treatment for their medical problems.	B10	0.845	0.87		

Table 5

Discriminant validity results

	Purpose		Benefit		Individualization		Any TM	
	t(154)=5.78, p<.0001		t(168)=3.66, p<.001		t(151)=4.59, p<.0000		t(185)=5.17, p<.0001	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
TM in interview	38.61	7.19	32.58	12.40	34.42	11.14	33.05	11.87
	n=28		n=66		n=45		n=95	
No TM in interview	25.53	11.47	25.93	10.94	25.19	11.43	22.63	9.75
	n=128		n=104		n=108		n=92	

Diagnostic accuracy measures

Table 6

Cutoff	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LR+ (95% CI)	LR− (95% CI)
Original results						
24	.79 (0.697 – 0.859)	.52 (0.421 – 0.621)	.63 (0.541 – 0.712)	.71 (0.589 – 0.801)	1.65 (1.302 – 2.093)	.40 (.261–.624)
27	.74 (0.640 – 0.815)	.61 (0.507 – 0.702)	.66 (0.566 – 0.744)	.69 (0.584 – 0.781)	1.88 (1.421 – 2.496)	.43 (.297–.629)
35	.52 (0.417 to 0.614)	.78 (0.688 to 0.855)	.71 (.594 – 0.804)	.61 (0.52 to 0.693)	2.37 (1.537 – 3.662)	.49 (.619–.782)
10-fold cross-validation results						
24	0.79 (0.697 – 0.859)	.52 (0.421 – 0.621)	.63 (0.541 – 0.712)	.71 (0.589 – 0.801)	1.65 (1.302 – 2.093)	.40 (.261–.624)
27	0.72 (0.618 – 0.797)	.61 (0.507 – 0.702)	.65 (0.558 – 0.738)	.68 (0.568 – 0.766)	1.89 (1.376 – 2.432)	.47 (.326 – .668)
35	0.55 (0.447 – 0.644)	.72(0.618 – 0.799)	.67 (0.556 – 0.761)	.61 (0.512 – 0.692)	1.94 (1.333 – 2.814)	.63 (.489 – .815)