

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES PUBLIC HEALTH SERVICE ALCOHOL, DRUG ABUSE, AND MENTAL HEALTH ADMINISTRATION NATIONAL INSTITUTE OF MENTAL HEALTH 試験評価手順スケール (Trial Assessment Procedure Scale)(TAPS) Jerome Levine	試験名	種類	治療法	状態	相	評価者
	評価者名				日付	
	所属					
	住所					
	電話番号					

試験名および/または識別番号 : _____
 報告書の種類、出典、および日付 : _____
 試験対象薬名又は試験対象治療法名 : _____
 状態 : 計画中 終了
 相 : 第 相 前期第 相 後期第 相 第 相 第 相
 治療群の数 : _____ 被験者数 : _____

解説

概観 試験評価手順スケール(以下 TAPS)とは臨床試験の品質評価を目的とした系統的な技法である。この技法は、試験の品質を反映する多くの記述的特徴または事項に関する報告書(例えば、プロトコル、終了後の試験報告書、または専門誌論文)を分析するものである。この事項は論理的に8つのカテゴリーにまとめられ、試験の様々な要素の質を独立して評価することができる。この目的は治療効果または安全性に関する知見に関係なく、試験の品質を格付けすることである。

各カテゴリーは2~5つの関連した事項で構成される。例えば、第1カテゴリーの研究課題は、「背景および根拠」と、「目的および/または仮説」という2つの事項から成る。事項カテゴリーごとに格付け用ページがあり、構成する事項が、それぞれの事項の評価において考慮すべき要素の種類とともに、一覧表に記載されている。これらの例は代表的なものであり、ある事項を格付けする際に考慮される事柄全てを網羅するものではない。

格付け方法 TAPS の格付け手順は全ての事項に対して同一である。格付けは5段階スケールで行われる:「十分満たす」、「満たす」、「境界」、「不十分」または「全く不十分」である。評価者は事項ごとに品質格付け欄の然るべき にチェックするよう求められ、評価中の試験がその事項に関してどの程度十分になされているかを反映する。評価者はいかなる事項の格付けに関してもその根拠の記載が奨励されており、このための余白欄が設けられている。

格付け上の問題点 臨床試験報告書の性質および内容における相違のため、ある与えられた事項を意味を持つように格付けすることが困難な場合がある。そこで格付け上の問題点および理由という見出の下に が2つ用意されている。最初の は「適用不能」で、これは試験の種類またはそれを通して試験を評価する報告書の種類（例えば、プロトコル、専門誌論文）のいずれかに対する事項の適用可能性のことである。この にチェックするのは、当該試験の脈絡または報告書の種類のどちらかにおいて、その事項が意味を成さない時のみであり、故に品質格付けを実施しない。2番目の、「記述不備」は、当該事項について意味ある判断を行うために必要とされる記載内容の有用性および明瞭性に関するものである：したがって、この がチェックされるべき時は、適切な情報が欠落しているまたは不十分かのどちらかである。たとえその事項について品質格付けを実施した後でも、この にチェックをし、関連情報が不十分であることを示しうる。

評価者は、以下のルールに留意されたい。評価者は、特定の分野の専門知識に自信がない場合でも格付けの実施にしり込みすべきではない。次に、ある事項の内容が試験自体または報告書の種類に適合しないために格付けが困難な場合、「適用不能」にチェックし余白欄に理由を記載する。その事項に関しては品質格付けを実施しない。（適用できたであろうはずの）当該事項に関し、判断が下せるだけの情報が報告書にないために、格付けが困難である場合「記述不備」にチェックする。与えられた情報が不明瞭、不完全、または推測しなければならない場合は、格付けを実施した上で、「記述不備」にチェックする。このように、事項の品質格付けを実施しない場合は必ず「適用不能」か「記述不備」のいずれかにチェックし、理由を記載しなければならない。品質格付けを実施したが、情報不足を評価者が示したい場合は、「記述不備」にチェックをしておく。

全体格付け 全ての事項に関する格付け終了後、全体格付けを実施し、総合的品質、あるいは試験全体の『よさ』に関する評価者の評価を示す。全体格付けには、記載されている他のいかなる検討事項とともに、事項カテゴリー全体に渡る個々の格付けすべてを考慮に入れるべきである。例えば、ある事項に関する品質格付けが、あまりにも受け入れがたいもので試験が『致命的欠陥有り』となる場合、たとえ他の事項の格付けが「満たす」以上と判定されていたとしても、全体格付けは非常に低くなる。全体格付けの箇所に表示される通り、格付けは0～100のスケールで行われ、0は「非常に悪い」、100は「非常に良い」を示す。0から100までのどの数値でも使用できる。全体格付けスケールの下に試験の総合的格付けについてコメントが記入できる欄がある。この欄はまた、試験、各格付け、および他の関連検討事項についての追加コメントにも使用できる。

試験報告書の評価 評価者は、TAPSの形式及び内容をよく理解したら、まず、格付け対象の試験報告書全体を読むことを奨励する。その後、必要に応じて試験報告書のさまざまな項を参照しながら、TAPSに記載されている順序で事項を格付けすべきである。

TAPSのスコアリング 格付け完了後、スコアリングの方法にしたがって一連の数値スコアが算出できる。

・ 研究課題

事 項	品質格付け	格付け上の問題点および理由	
A. 背景および根拠 - 以前の関連研究の知見を適切に提示している；研究の必要性および検証される仮説の基礎 / 根拠の正当性。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
B. 目的および / または仮説 - 研究課題の目的、重要性、および精度の明瞭性；なされるべき主張に対する仮説の関連性。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備

コメント：

・ 研究管理

事 項	品質格付け	格付け上の問題点および理由	
A. 外部 審査 / モニタリング - 独立した有資格グループによる科学的および倫理的な審査の妥当性；研究の運営、状態、および進行に関する外部モニタリングの使用。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
B. 施設の選択 - 臨床施設の選択の明瞭性と客観性；治療および評価状況の妥当性。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
C. 人員 - スタッフの組織構成の適切性、例えば監督の適切性；患者のケアの実施、評価、データ解析に関する個々のスタッフの専門技術。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
D. 試験期間 - 計画段階、データ収集期間および解析期間の長さの適切性、およびデータ収集完了から解析開始までの間隔の適切性。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備

コメント：

. デザインの特徴 事 項	品質格付け	格付け上の問題点および理由	
A. 独立変数 - デザインに関わる要素の選択、例えば、治療法、治療内での薬物の選択（実験薬、標準薬、プラセボ）、患者の診断、投与期間、等々。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
B. デザイン構成 - 試験的デザインの妥当性と正確性、例えば、必要に応じたクロスオーバーデザイン、または独立グループデザインの使用など；例えば、治療順序、時間、設定などの治療因子と交絡しうる影響の回避。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
C. 被験者の割り付け - 検証する仮説に対する、サンプルサイズ（各治療群内）の妥当性；適切なランダム化、マッチング、連続手法などによる、被験者の適切な治療グループへの割り付け。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
D. 治療関連バイアスの管理 - 治療および評価者盲検化（例えば、二重または三重盲検）の適切性；投与スケジュール、投薬剤型、投与時間の比較可能性；患者全員への盲検は解除せずに、個々の患者への盲検だけを解除する対応策；被験者の疾患の著しい改善または悪化、または治療による毒性副作用の発現に対処するための明確な規定または規準の利用。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
E. 外部変数の管理 - 研究環境の適性、例えば、著しい調査者バイアス、“ホーソン効果”、あきらめの雰囲気などが存在しない；例えば、層別の使用、持ち	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備

越し効果の回避などによる治療前バイアスの軽減；研究で使用する薬物以外の薬物を使用する他の併用療法の制限または管理；併発疾患、引越しなどの問題の扱いに対する明確な規定または規準の利用。

コメント：

・ 治療の特徴

事 項	品質格付け	格付け上の問題点および理由	
A. 記述 - 研究対象治療および対照治療に関連した特徴、例えば、推定される臨床作用、副作用、作用持続期間、および薬理学的作用などの明確性；比較対象薬剤、すなわち、標準薬またはプラセボ、の選択の根拠。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
B. 投与量 - 用量レベルの妥当性、標準および試験薬の用量等価性、用量調整に関する規準；作用持続時間、研究デザイン、試験のフェーズに関して、投与スケジュールおよび投与パターン（固定または可変）の適切性；投与剤型または投与方法の適切性、薬理学的性質の一致度合。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
C. 期間 - ドライイング・アウト（治療前）期間、薬剤投与（治療）期間、および追跡（治療後）期間の必要性および長さの妥当性。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備

コメント：

・ 被験者の特徴

事 項	品質格付け	格付け上の問題点および理由	
A. 選択規準 - 患者を診断して試験に患者を組み入れるまたは除外するために使用する規準の明瞭性、的確性、妥当性、および一般的容認。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
B. サンプルの代表性 - 疾患に関連した特徴（例えば、精神病理学のパターンと重症度）および人口動態学的 / 状況関連特性（例えば年齢、性別、疾患の急性度、入院 / 外来設定など）の点で、サンプルと母集団の間の相関性。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
C. 被験者の組み入れ - 被験者のリクルート方法の適切性と一貫性、および自発的被験者のインフォームド・コンセントを得るための必要条件に対する遵守度合い。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備
D. 被験者のコンプライアンス - 評価方法やスケジュールへのコンプライアンスおよび薬物摂取に対するチェックおよび確認技術の適切性。	十分に満たす 満たす 境界 不十分 全く不十分	適用不能	記述不備

コメント：

・ データ収集

事 項	品質格付け	格付け上の問題点および理由
<p>A. 評価の範囲</p> <p>- 以下のような評価対象分野の測定範囲の妥当性：例えば、サンプル（疾患の程度、人口動態学的情報などの固定および記録）、効力（患者の疾患の改善または悪化を示す評価）、副作用（予想されたまたは予想されない治療誘発性症状を診断する評価）、投与量（実際に投与された用量の記録）、安全性（試験対象薬物に特異的かつ適切な臨床検査の使用、身体機能の評価全般、など）、他の関連分野（疾患、薬物、または特殊な評価技術に依存する分野、例えば EEG、血中濃度、行動評価など）。</p>	<p>十分に満たす 満たす 境界 不十分 全く不十分</p>	<p>適用不能 記述不備</p>
<p>B. 評価方法</p> <p>- 評価対象分野に関し、選択された設定方法および機器の妥当性；評価スケールおよび記録様式の、以前に示された感度、信頼性、妥当性の程度；測定方法の一般的な使用、容認度合い。</p>	<p>十分に満たす 満たす 境界 不十分 全く不十分</p>	<p>適用不能 記述不備</p>
<p>C. 評価スケジュール</p> <p>- 評価実施の頻度およびスケジュールの適切性、試験開始前のベースライン測定値の収集。</p>	<p>十分に満たす 満たす 境界 不十分 全く不十分</p>	<p>適用不能 記述不備</p>
<p>D. 評価の実施</p> <p>- 試験全体を通じて、評価実施および評価技術の適用の一貫性；各患者に対して試験全体を通じて同一の評価者を当てる試み。評価者間の信頼性と当該試験事情における評価妥当性の確立の証拠。</p>	<p>十分に満たす 満たす 境界 不十分 全く不十分</p>	<p>適用不能 記述不備</p>

コメント：

．データ解析

事 項	品質格付け	格付け上の問題点および理由	
<p>A. データの準備</p> <p>- データ収集技術の適切性（症例報告書、記録様式など）；データチェック、編集、および証合技術の適切性；手計算か、それとも標準的コンピュータ解析プログラムを使用するか、中間データ処理段階でのチェック。</p>	<p>十分に満たす 満たす 境界 不十分 全く不十分</p>	適用不能	記述不備
<p>B. データ表示</p> <p>- データ記述、構成および表示の明瞭性、重要性および利用性；データの細部または要約の適切なレベル。</p>	<p>十分に満たす 満たす 境界 不十分 全く不十分</p>	適用不能	記述不備
<p>C. 統計解析</p> <p>- 解析手法適用の正確さ、例えば、データ変換、データ集積、脱落および欠落データの処理など；研究デザインに適した統計モデルの使用、例えば、パラメトリックかノンパラメトリックか、分散分析か共分散分析かなど；統計的な最終仕上げ、例えば、有意な F 比判明後の多重比較の使用など。</p>	<p>十分に満たす 満たす 境界 不十分 全く不十分</p>	適用不能	記述不備
<p>D. データの統合</p> <p>- 研究課題に答えるための系統的アプローチの実践；関数関係の適切な解析（例えば、用量 反応曲線）。</p>	<p>十分に満たす 満たす 境界 不十分 全く不十分</p>	適用不能	記述不備

コメント：

・ 結論と解釈

事 項	品質格付け	格付け上の問題点および理由	
A. 焦点 - 提示された結論（知見）と解釈（説明）が特異的で明確である程度；結論と研究仮説間の重要な一致。	十分に満たす	適用不能	記述不備
	満たす		
	境界		
	不十分		
	全く不十分		
B. 論理 - 試験において収集されたデータにより、理論的かつ統計的に、結論が明白に支持される度合い。	十分に満たす	適用不能	記述不備
	満たす		
	境界		
	不十分		
	全く不十分		
C. 適用 - 研究標本からより大きな母集団への、結果一般化の妥当性（即ち、主張は支持データ以上の誇張または過剰な一般化をしていない）；公表された他の知見の解釈の適切性。	十分に満たす	適用不能	記述不備
	満たす		
	境界		
	不十分		
	全く不十分		

コメント：

試験の総合的品質



左のスケールに従い、
0 から 100 までの数値を記入する。

全体格付け

コメント：

スコアリングの方法

本ページでは、TAPS 格付けからスコアを割り出すための方法と余白を設けてある。スコアリングの方法は以下の通りである。

カテゴリーのスコア このスコアはすでに記録済の事項「品質格付け」から割り出す。

(1) カテゴリー毎(~)に事項の「品質格付け」を以下の対応する事項数値スコアに換算する：

十分に満たす = 100 点； 満たす = 75 点； 境界 = 50 点； 不十分 = 25 点； 全く不十分 = 0 点。これらのスコアを下記の表の該当箇所に記録する。

(2) カテゴリー毎の事項スコアを合計し表に記入する。カテゴリー毎に格付けされた事項の数を記入し、その事項スコアの合計を格付けされた事項の数で割り、その結果であるカテゴリースコアを表の指定欄に記録する。(あるカテゴリーでどの事項も格付けされなかった場合、カテゴリースコア無しとなる)

事 項	カテゴリー							
	_____	_____	_____	_____	_____	_____	_____	_____
A	_____	_____	_____	_____	_____	_____	_____	_____
B	_____	_____	_____	_____	_____	_____	_____	_____
C		_____	_____	_____	_____	_____	_____	_____
D		_____	_____		_____	_____	_____	
E			_____					
事項スコアの合計	_____	_____	_____	_____	_____	_____	_____	_____
格付けされた事項の数	_____	_____	_____	_____	_____	_____	_____	_____
カテゴリースコア	_____	_____	_____	_____	_____	_____	_____	_____

総スコア 算出された全カテゴリースコアを合計し、カテゴリースコアの数(通常 8 個)で割る。

カテゴリースコアの合計 _____ ÷ カテゴリーの数 _____ = 総スコア _____

スコア差 全体格付けと総スコアの数値差(即ち、絶対値)を算出する。

スコア差を記録する。 _____

記述不備のスコア 記述不備の _____ にチェックが入っている事項の数を合計し記載する。 _____