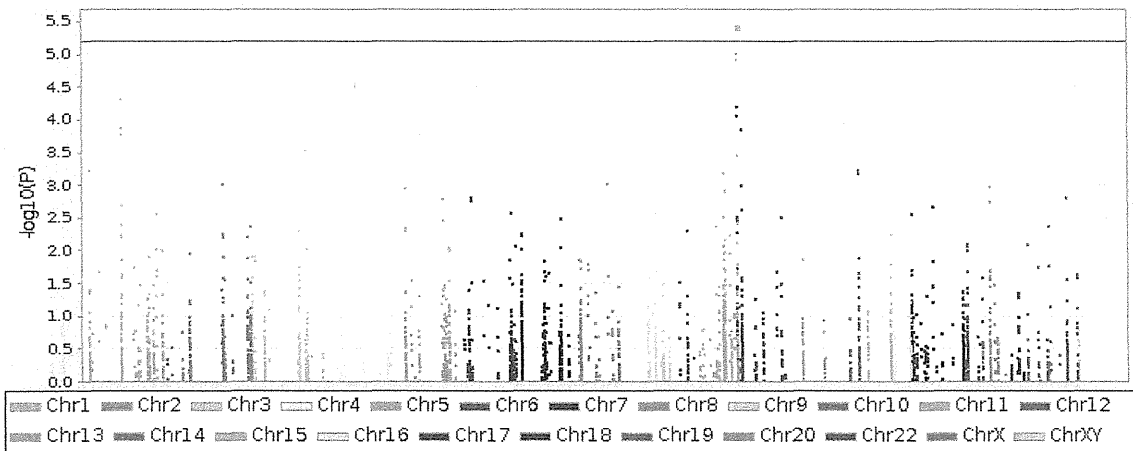
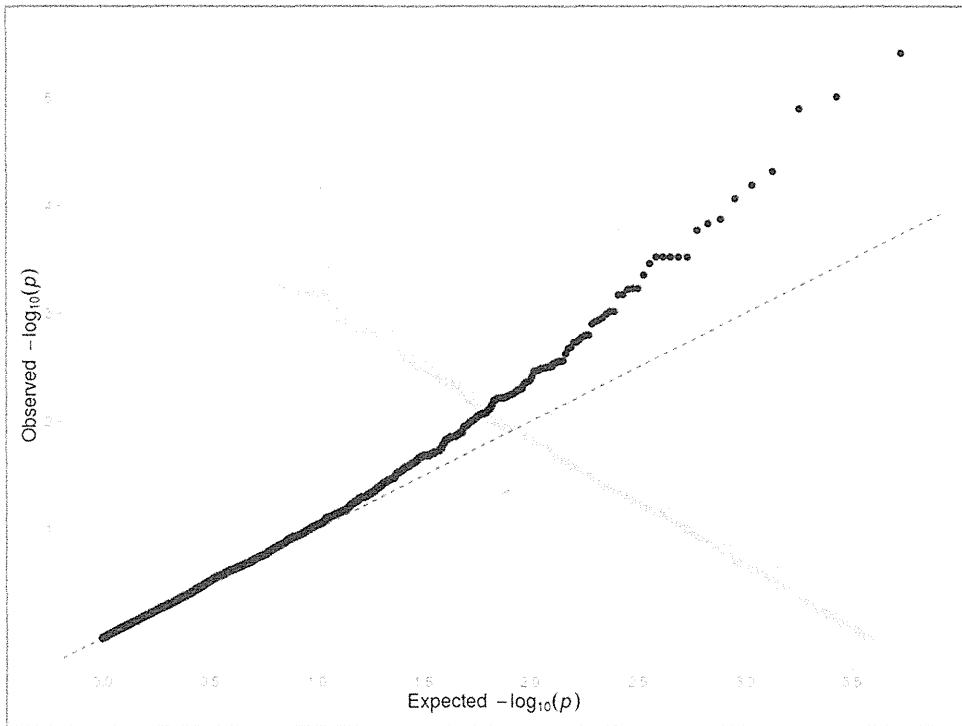
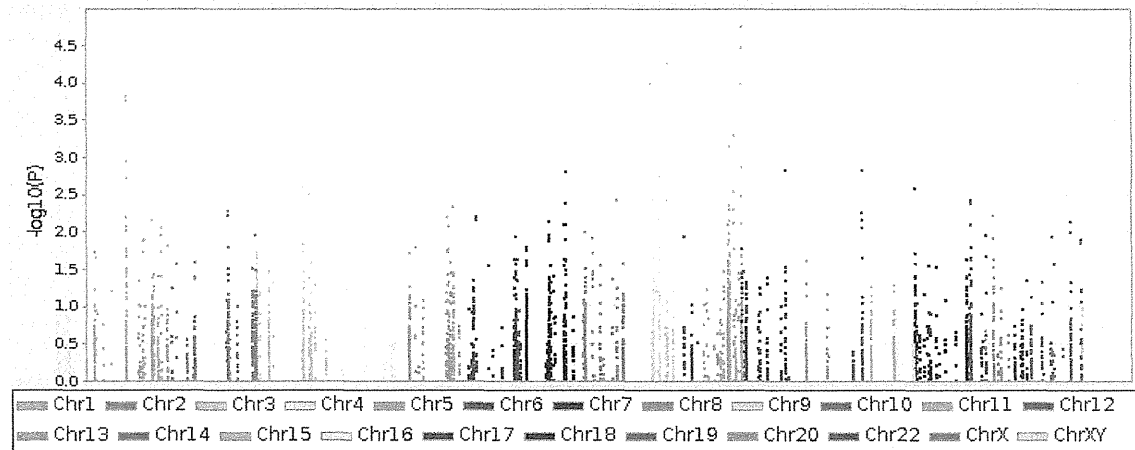
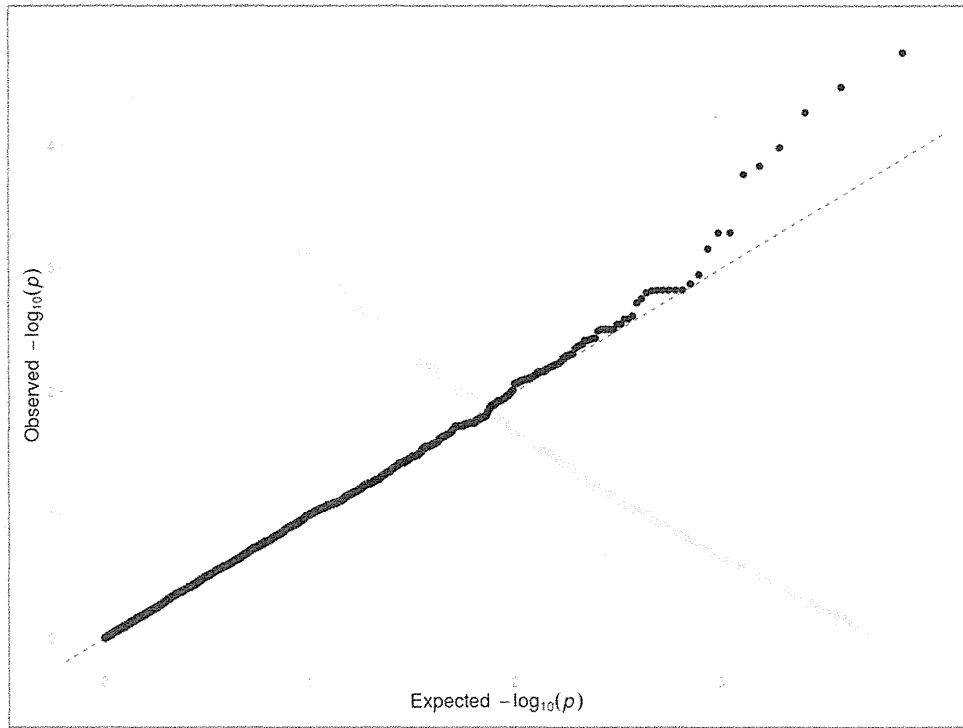


遺伝型様式



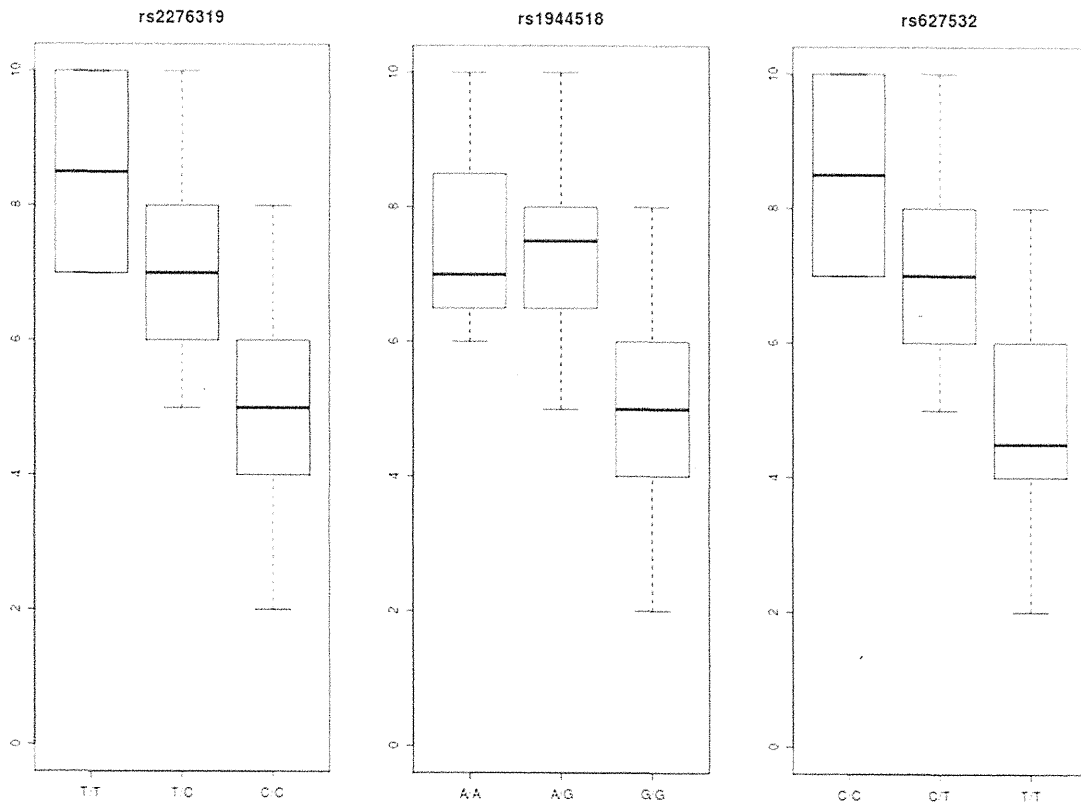
傾向性様式



優性様式、劣性様式、傾向性様式において関連性を示している SNP はありませんでした。遺伝型様式では BH 法で有意と判断された SNP が 3 つ、ボンフェローニの補正法によって有意と判断された SNP が 1 つ存在しました。有意となった SNP の情報は以下の通りです。

SNP	CHR	Gene	Location	Major Allele	Minor Allele	MAF	P	Bonf.Sig	BH.sig
rs2276319	11	GRIK4	120343031	C	T	0.26	3.85E-06	Yes	Yes
rs1944518	11	GRIK4	120348952	G	A	0.27	9.71E-06		Yes
rs627532	11	GRIK4	120331341	T	C	0.26	1.26E-05		Yes

SNP		rs2276319	rs1944518	rs627532
Gene		GRIK4	GRIK4	GRIK4
Allele	Major	C	G	T
	Minor	T	A	C
Phenotype	MajorHomo	5 ± 1.64(N=22)	5 ± 1.64(N=22)	5 ± 1.69(N=21)
	Hetero	7.41 ± 1.58(N=17)	7.5 ± 1.59(N=16)	7.28 ± 1.64(N=18)
	MinorHomo	8.5 ± 2.12(N=2)	7.67 ± 2.08(N=3)	8.5 ± 2.12(N=21)
Kruskal-wallis test		4.17.E-04	4.85.E-04	8.77.E-04



【Pain 2010; 149: 177-93】

## Research design considerations for confirmatory chronic pain clinical trials: IMMPACT recommendations

Robert H. Dworkin<sup>6\*</sup>, et al.

### 抄録

慢性疼痛への有効性が期待できることを示唆する過去の研究があるにもかかわらず、主要有効性解析で評価対象の治療とプラセボとの間で有意差がないことを示す慢性疼痛に関する臨床試験の数が増えている。これらの所見は、有効性の真の欠如あるいは有効性の証明に支障をきたす方法論的な面なども反映している可能性がある。慢性疼痛に関する臨床試験の間で重大な研究デザインの検討事項にかなりのばらつきがあることから、方法論的な弱点を明らかにし対処することによって新たな介入の鎮痛作用を実証する可能性が強まるであろう。このような背景から、慢性疼痛に関する確認試験の研究デザインの重要な検討事項を特定し、これらの試験実施に関する勧告を行うことを目的としたIMMPACT コンセンサス会議が召集された。本稿では、試験の種類に加え、参加者の選択、試験相および試験期間、治療群および投与レジメンなどの慢性疼痛に関する確認臨床試験の主要要素に対する推奨事項を提示する。慢性疼痛に関する確認臨床試験の方法論的な面に注目し、研究を増やすことにより分析感度が高まり、最終的に慢性疼痛治療に関する有意義な評価を提供できると思われる。

キーワード: Chronic pain (慢性疼痛)、Randomized clinical trials (無作為化臨床試験)、Research design (研究デザイン)、Phase 3 trials (第 III 相試験)、Subject selection (被験者の選択)

### 1. 緒言

慢性疼痛治療に関する臨床試験の実施および解釈を促すために、Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) は、成人[29,143]、小児、青年[95]に関する一連のコア・アウトカム領域と指標に加え、改良アウトカム指標の作成[144]、個々の患者の改善および群間差の臨床的重要性の判定[30,32]、ならびに複数アウトカムの分析[145]の手法を推奨している。これらの推奨事項は、慢性疼痛試験でのアウトカム指標の選択、分析、解釈を含んでいるが、同じように重要な慢性疼痛に関する臨床試験で使用する研究デザインについての検討項目はない。鎮痛薬に関する臨床試験の試験デザインに大きな影響力のあった論文がほぼ 20 年前に発表されているが[94]、その後、慢性疼痛に関する無作為化臨床試験 (RCT) の研究デザインに大きな進展がみられる。

二重盲検プラセボ対照 RCT は依然として、有効な治療介入を特定するための広く受け入れられている唯一の研究デザインであるが、慢性疼痛などの複雑な病態の検討に際しては大きな限界がある[62]。このデザインは、慢性疼痛を軽減するために設計された新規の介入について規制当局に承認を申請する際のピボタル試験で典型的に用いられている。しかし、そのような慢性疼痛試験で用いられる研究デザインおよび方法が注目されるようになったのは、最近のことである[27,31,68-70,107,108]。これは、主として臨床試験の数が増加した結果であるが、先行する前臨床研究や臨床研究で有効性が期待されることが示唆されているにもかかわらず[14,16,103,132,133,140,148,153 など]、これらの臨床試験で評価対象薬剤は主要有効性解析においてプラセボとの間で有意差を認めなかった。

これらのネガティブな所見が一般的条件または特定条件の下で有効性が真に欠如していることを反映しているかどうか、他の因子(わずかな治療ベネフィットの検出力が不十分、プラセボ群の奏効率が高過ぎる、データが欠落し補完を必要とした脱落例が多い、など)が有効性の証明を難しくしたかどうか、あるいはその両方が関与していたかどうかは明らかでない。慢性疼痛に関するRCTの研究デザインにはかなりのばらつきがあるため、これらの試験の方法論的特徴が有効な治療のベネフィットの証明に関与していた可能性もある。これらの結果の一部が研究の失敗を反映しているとする仮定は、慢性疼痛についての 2 件以上の試験で有効性が証明されていた薬剤に関する最近のネガティブな RCT によって裏付けられている[153]。複雑な慢性疼痛の病態も、臨床試験から得られる情報の種類に限界があるようである[62]。重要なことは、真に有効な介入に関するネガティブな試験は、既存治療に難治性の患者に有用かもしれない治療の特定を妨げるだけでなく、科学的・金銭的なリソースもかなり消費される点である。

これらの理由から、研究デザインの改善が臨床的に有用な治療の特定能の強化につながることを期待して、慢性疼痛治療に関する RCT の研究デザインにおける重要な検討事項を特定し、これらの試験実施に関する勧告を行うことを目的とした IMMPACT のコンセンサス会議が召集された。本稿では、「確認(confirmatory)」または「第 III 相(Phase 3)」と称した試験、すなわち、最低でも有効性および安全性に対する予備的エビデンスが既に存在する治療の有効性および安全性について、確かなエビデンスを得るために実施された適切な対照群のある(well-controlled)試験に注目した。有効性についての最初のエビデンスが得られる臨床試験デザイン(すなわち、「概念実証[Proof-of-concept]」または第 II 相[Phase 2]試験)に対する推奨事項は次回の IMMPACT 会議のテーマであり、別に報告す

る予定である。

本稿では、慢性疼痛に関する確証臨床試験のための研究デザインの検討事項を示すことを目的とし、以下の主要項目を含む：(1) 研究被験者、(2) 試験相、(3) 治療群、(4) 確認試験の種類。慢性疼痛に関する臨床試験の分析および解釈に関連した種々の事項については、他で報告されている[30,32,145]。本稿の基となった IMMPACT 会議では薬物療法の確認試験に注目しているが、多くの検討事項は他の慢性疼痛治療(理学療法、脊髄刺激、心理社会的介入など)に関する試験や異なる治療法の併用に関する研究などの臨床試験も含んでいる[78]。頭痛に関する臨床試験の研究デザインは、本稿の範囲外である[131]。

## 2. コンセンサス会議の手順

IMMPACT コンセンサス会議は、大学、政府機関、産業界、患者擁護団体からの参加者による国際的グループが参加して開催された。参加者は、慢性疼痛治療に関する試験のデザインおよび評価に関する研究、臨床上または行政上の専門性に基づいて選択された。率直で効率的な議論を推進するために会議の規模を制限し、種々の学問分野および専門性を幅広く代表する人々を含めることとした。議論およびコンセンサスに基づく勧告の開発を促すために、会議では補足的講義で以下が検討された：(1) 上乘せデザインに焦点を合わせたてんかんに関する臨床試験(JAF)、(2) プラセボ群の反応に焦点を合わせた抑うつに関する臨床試験(Michael E. Thase, MD)、(3) 臨床試験での効果の発現および持続性の分析および解釈(SE)、(4) 強化した組入れによる無作為化治療中止試験(NPK)。以下の推奨事項の多くは系統的な研究に基づいたものではなく、むしろ慢性疼痛やその他の疾患(抑うつ、てんかん、など)の治療に関する臨床試験の方法および結果からの外挿であることを強調することは重要である。

## 3. 研究対象

臨床試験の組入れ基準および除外基準は参加者の特徴を設定するものであることから、広範囲な集団に対する一般化に影響を及ぼすだけでなく、試験が有効性を首尾よく証明できる可能性にも重大な影響を及ぼす[115,116]。これらの 2 つの目標はしばしば対立する。例えば、地域社会で患者の不均一性を反映した状態の対象者を採用することにより試験の治療効果の検出力が低下する可能性がある。例えば、患者の人口統計学的特徴および疾患重症度、罹病期間、併存疾患、併用薬などの臨床的特徴に大きなばらつきがあれば、治療に対する患者の反応にも大きなばらつきが生じる。組入れ基準の設定では、臨床試験の「分析感度」を損なう可能性があるばらつきの既知の原因や潜在的原因を有する幅広い集団を反映した対象者の特徴に慎重なバランスを取ること、すなわち、試験がプラセボまたは実薬対照との比較で治療効果を証明する能力が要求される。

組入れ基準が極めて具体的な(例えば、HIV に関係した神経障害)確認試験で実薬治療の有効性を示した結果は、臨床診療の場でこの治療法が有効でないとと思われる患者にも(例えば、腰痛神経根障害)、この治療が行われる危険性がある。逆に、幅広い患者集団(例えば、あらゆる種類の神経障害性疼痛)を検討した試験での有効性欠如は、少なくとも一部の患者に有効であると思われる治療が放棄される危険性がある。この点においては、例えば、極めて特異的な病態に対する新たな化

合物の有効性を証明する「説明的」試験と、例えば、よくある不均一性の臨床症状(例えば、腰痛)に対する既知治療の有効性を評価する「実践的」試験の区別が役に立つ[127]。このように、臨床試験は予め慎重に設定された目的に基づいて治療を評価するように設計されるべきである(例えば、特定の適応に関する規制当局の承認とプライマリケアにおける有効性の推定を区別する必要がある)。

### 3.1. 人口統計学的特徴、その他の特徴

すべての臨床試験に関する出版物およびオンライン登録は、全被験者について少なくとも年齢、性別、人種/民族に関して報告する必要がある。これらの特徴の中で年齢のみが組入れ基準の 1 つとして頻繁に使用されている。例えば、18 歳未満の患者は初期の確認試験から除外されることが通例であり、帯状疱疹後神経痛(PHN)のような小児や青年が罹患することがほとんどない特定疾患に関する臨床試験では除外できる。さらに、高齢患者(85 歳超)の場合、併存疾患の有病率や薬物療法の施行率が高い人に安全性や忍容性の懸念があるならば、試験対象から除外されることがある。人口統計学的データを広範囲に報告することで、治療の有効性、安全性、忍容性が患者背景に応じて変化するかどうかについての研究が可能となり、そのような分析はどの患者がその治療に適応となるかを証明する重要な要素となる。

その他の社会経済的特徴、心理社会的特徴、生活の質に関する特徴も慢性疼痛患者の場合、特に重要となることがある。例えば、被験者の教育や職業の状況ならびに健康関連の生活の質は、薬物療法および非薬物療法に関する臨床試験において治療反応と関連していた[例 67,100]。さらに、疼痛に関係した現在訴訟中の患者や労働災害補償を受けている患者については、文書で証明されている疼痛の軽減が法的な状況または身体障害の状態に不利な影響を及ぼす可能性があり、それらの患者の疼痛改善の報告が妨げられることが懸念されるため、慢性疼痛に関する臨床試験から除外されることが多い。しかし、身体障害は慢性疼痛患者によくみられ、これらの患者を除外することは臨床診療の場での一般化に制限を加えることになるかもしれない。

治療アウトカムのばらつきの一因となる可能性がある既知の患者背景および潜在的な患者背景の徹底的な評価を臨床試験の開始に先立って実施すべきであり、これらは対象者を無作為割り付けする前に評価される必要がある[62]。一部の患者背景は疾患の進行や薬剤の代謝に影響を及ぼし、それによって治療反応にも影響を及ぼすことがある。例えば、体重は痛性糖尿病性末梢神経障害(DPN)に影響を及ぼす可能性があり、腎機能低下は腎臓で代謝される薬剤の副作用を増加させる可能性がある。患者を代謝が遅い人と速い人に分類可能な場合もあり[134]、薬物反応の可能性が最も高い患者を特定するために有効性および安全性に関する分析でこれらの情報の検討が可能である。ヒトゲノムに関する知見が増えるにつれ、RCT で遺伝子型解析および有効性が検証されているバイオマーカーを使用することにより、薬剤と被験者を薬物動態学および薬力学的な相互作用の非常に幅広い知識に基づいてマッチングさせることがさらに可能になると思われる。

### 3.2. 診断

正確な診断は、臨床試験に組入れられる被験者が治療対象疾患または状態を有しているかどうかを確認するために欠かせない。慢性疼痛症状はしばしば、このプロセスにおいて次の 2 つの要素が

ある：(1) 例えば、変形性関節症[OA]や糖尿病などの基礎疾患の診断、(2) 例えば、OA による膝関節痛や有痛性 DPN など疾患に起因する疼痛症状の診断。慢性疼痛を伴う疾患の多くは、広く受け入れられている診断基準がある。例えば、股関節の OA に対する米国リウマチ学会の基準では、股関節痛に加えて 2 項目以上の臨床検査による陽性結果および X 線上的所見が要求される[1,2]。

しかし、慢性疼痛症状の診断は時にきわめて難しい場合がある。最近改訂された神経障害性疼痛の定義は、2 種類の確認検査がある場合に限り「確定」診断が行われる評点方式がある[141]。神経障害性疼痛の原因となっている病態のいくつかに対して多様な診断アプローチが開発されているが[例 21,26,36,54,65]、確認検査ならびにその陽性および陰性予測の妥当性に関するエビデンスの入手は限られているようである。

このような妨害はあるが、病因的に異なるいくつかの疼痛症状については一般に広く受け入れられている診断基準が開発されており、多様な作用機序を有する治療の有効性を証明する際の基準となっている。これらの症状には、OA、関節リウマチ、DPN、PHN および線維筋痛症(FM)に関連した慢性疼痛がある。慢性腰痛は一般集団で非常に高頻度に見られるが、その診断基準はまちまちであり、神経障害性または非神経障害性の疼痛機序が治療の標的であるかどうか依存している。

特定の病因を標的とした治療に関する相次ぐ臨床開発は多くの場合、既存治療の有効性の証明に先立って作用機序や治療対象の根底にある病態が明らかにされている所見と一致している。そのため、早期の臨床試験が特定の慢性疼痛の症状に有効であるというエビデンスを示している場合、同一症状を対象として確認試験を実施することは妥当である。

診断的に均一性の高いサンプルでは有効性が証明できる可能性が高いと一般に考えられているが、これは不均一なサンプルは治療に反応しない症状のある対象者を含んでいる可能性が高いためである。従って、可能な限り診断手順および検査を確認試験で使用する前に標的集団においてそれらの精度を評価すべきであり、また明らかに標的症候を有する患者のみを組入れ、標的症候が存在する可能性しかない患者を除外できるような正確な診断プロセスであることが理想的である[122]。このことによって、試験結果が適用される集団の正確な特徴付けがさらに可能となる。

診断プロセスおよび関連手順を試験実施施設および試験責任医師の間で標準化することにより、より均一性の高い患者サンプルの確保が期待できる。診断が主として病歴や症状に基づく場合は、一貫性のあるアプローチを用いて医師は患者に問診を行い、標準化されたプロセスを用いてこの情報を捕捉すべきである。例えば、スクリーニングでの来院時に、症状を明らかにするために一連の問診票の使用が可能であるが、これらは全ての患者に対して同一の順番で示し、同一の文言を用いる必要がある。Patient-Reported Outcomes Measurement Information System (PROMIS)などの現在進行中のイニシアチブでは、患者の自己報告による疼痛および関連アウトカムを評価するために標準化されたアプローチを開発中であり、これによって、臨床試験との間で、また場合によっては異なる疼痛症状の間で評価の同等性が裏付けられると考えられる[18,110]。可能であれば、観察者間および観察者内の再現性および信頼性を前もって判定し、ばらつきの原因を少なくするための対策を講じるべきである。これらと同じような配慮を理学的診察および徴候の文書化の標準化でも適用すべきである。また、電気生理学的所見や X 線所見[例 2,90]などのさまざまな診断手順およびその結果の解釈についても、このような客観性が高いとされる手順であっても信頼できない可能性もあることから[56]、

これらの標準化にも適用すべきである。

病因に基づく診断が現時点では、慢性疼痛に関する RCT において標準とされているが、慢性疼痛の診断、評価、治療に対する至適アプローチは、まず患者の疼痛が説明できる病態生理学的機序を明らかにし、これらの機序を至適治療アプローチ選択の根拠として使用することが提案されている[9,92,151,152]。しかし、慢性疼痛に関する確認試験の設計において依然として、これらの機序を十分に明らかにすることは出来ず、またどの治療がそれらの機序を標的とするかについても明確にできない。現在進行中のトランスレーショナル・リサーチ(translational research)は、慢性疼痛管理への機序に基づくアプローチの基礎となると考えられる。このことは、臨床試験に疼痛機序を特定するための系統的評価を含む必要があると思われる[例 10,112,113,126]。そのような機序に基づく治療へのアプローチが開発され、効果が検証されるまでは、現時点では根底にある機序(例 中枢性感作)よりも疾患に特異的な診断(例 膝関節 OA)を標的とした試験を続けている。

### 3.3. 罹病期間および疼痛持続期間

慢性疼痛試験では、研究対象者の慢性疼痛が自然に消失する可能性がある比較的軽いものであるかどうかを確認するために罹病期間の評価が必要である。慢性疼痛治療の有効性評価に際しては、疾患のばらつきに伴う変動を最小限に抑えるために基礎疾患の状態が安定している患者が一般に望ましい。しかし、特定の病態は重要な時間的変化を特徴としており(例 多発性硬化症)、これは試験デザインで対処しなければならない。典型的なものとして、慢性疼痛に関する確認 RCT の組入れ基準は、基礎疾患の診断からの最短期間および慢性疼痛の最短罹病期間を必要としている。例えば、有痛性 DPN に関する臨床試験では、対象者は DPN の診断を 1 年以上前に受けており、DPN に関連した疼痛を 6 ヶ月以上経験していることが必要とされる場合がある。

疾患の発症を知れば、慢性疼痛が特定状況の下で存在するかどうかを十分に確認できるかもしれない。例えば、帯状疱疹の発疹の発現日から、PHN の臨床試験で疼痛持続期間が簡単に判定できる[26]。しかし、他の多くの病態については、疾患持続期間は症状持続に関する患者の報告に基づいて確認しなければならない。疼痛持続期間に関する患者の報告は組入れ基準が満たされているかどうかを確認するために用いられることが多いが、患者によるこれらの報告の収集は必ずしも十分に開発された手法を用いて系統的に行われていない。診断コードまたは明らかな前治療歴(例えば、試験開始前の特定期間の鎮痛薬処方[125])を用いて症状持続期間に関する患者の報告を補強することにより、患者の回想の正当性を立証できるかもしれない。しかし、これらのデータは必ずしも入手できないことから、このような組入れ基準は選択バイアスの原因となるかもしれない。アプローチにかかわらず、臨床試験に関する出版物は期間判定に用いられた具体的な方法を明確に記述すべきである。

3 ヶ月以上持続している疼痛は慢性とみなされ[97]、多くの臨床試験でこの期間が組入れ基準の 1 つとされている。しかし、6 ヶ月以上の疼痛持続を必要とすることによって、自然経過の結果、疼痛の著しい軽減を経験する可能性がやや低い対象者サンプルが得られるであろう。このように厳しい疼痛持続期間を用いると組入れスピードを遅らせることにもなるため、既存のデータベースを検討して疼痛持続期間が比較的短い被験者を解析対象から除外した場合に分析感度が高まるかどうかを判定す

る必要がある。

個別のカットオフ・ポイントについては任意に設定されているが、5～10年の罹病期間または疼痛持続期間の上限[例 51,108]も慢性疼痛に関する一部の試験で用いられている。このような疼痛持続期間の上限を用いることで、前治療の効果がみられない難治性の患者や治療効果があるとは思えない非常に長い疼痛持続期間を有する患者の除外に有用であると思われる。しかしながら、治療が不十分であった患者や既存の療法と比較して異なる作用機序のある新たな介入に反応すると考えられる患者が除外される可能性もある。データが欠落している場合には、疼痛症状の病態生理学的機序、既存の治療と比較して異なる治療の作用機序、十分に確立された疼痛患者の多くの除外が組み入れに及ぼす影響を慎重に検討した上で、罹病期間または疼痛持続期間の上限を慢性疼痛試験に組み入れるべきである。

### 3.4. 疼痛強度

総合的または平均的な疼痛強度は、同一の疼痛症状を有する患者の間でもかなりのばらつきがある。このばらつきには、疼痛感受性に対する個人差[34,41]、活動レベル、対処能力、基礎疾患または疼痛原因の重症度など複数の原因がある。現在のところ、慢性疼痛に関するほぼ全ての試験で、ベースライン時の患者の自己報告による疼痛強度が一定以上(多くの場合、0～10の数値化スケール[NRS]で4)であることを要求している[7,90,118]。一般に、このような測定尺度の心理測定的特徴は、尺度の中央の最初のスコアが測定の天井および床効果の低減によって変化に対する感度を改善させる。しかし、慢性疼痛に関する一部の試験では、疼痛強度の最低ラインを低く設定するか高く設定するかいずれかであり、例えば、0～10のNRSで3[43]または5[25]と設定されている。個々の被験者の疼痛強度の最低ラインを4に設定すると、ほとんどの場合、神経障害性疼痛に関する試験では無作為割り付けされた患者のベースラインの群平均疼痛強度は5～7となるが[68]、OAに関する一部の試験では7～8の平均疼痛強度が観察されている[6]。

欧州医薬品庁[38]は、軽度の疼痛患者ではプラセボ反応率が高くなる可能性があるとの理論的根拠[参照 61,155]に基づき、神経障害性疼痛に関する臨床試験では中等度以上の疼痛を有する患者を組入れることを推奨している。大うつ病に関する臨床試験では、高い抑うつ重症度は治療効果が大きいことと関連していることを示すある程度のエビデンスがあるが[27,77,82]、ベースライン時の疼痛レベルが高ければ高いほど、治療効果が大きいかどうかについて検証した慢性疼痛試験は比較的少ない。有痛性DPNに関する3件の試験を対象とした最近の分析では、Zieglerら[157]は、ベースライン時の疼痛強度が0～10のNRSで6以上の被験者と無作為割り付けの最低要件であった4以上6未満の被験者の比較から、治療の絶対差が大きいことを指摘している。OAまたは関節リウマチでベースライン時の疼痛強度が高い患者は、ベースライン時の疼痛強度が低い患者よりも1クールの実薬治療による改善が大きいことも示されているが[11,12,15]、ベースライン時の疼痛強度が高い患者でプラセボまたは実薬対照との比較で高い有効性が認められるかどうかは不明である。

重要なことは、臨床試験の組み入れ基準が限られている場合(すなわち、要求される標的症状のレベルが高い)、患者が組み入れ基準を満たすようにベースライン時の症状重症度の評価の「インプレッション」が起こる可能性があることを示すある程度のエビデンスがある[79,82]。そのようなベースライン時の

インプレッションが研究実施施設の財政面などの検討事項や患者が新規治療を受ける機会を得たいという動機付けなどの因子を、どの程度反映しているかどうかは不明である。しかしながら、無作為割り付けに求められるベースライン時の疼痛強度を高く設定することによって慢性疼痛試験の分析感度を高めようとする全ての試みには、ベースライン時の疼痛強度の潜在的インプレッションを伴う可能性があり、このことによって分析感度が低下する可能性がある。慢性疼痛の臨床試験で、ベースライン時の疼痛の意図的または非意図的なインプレッションが起こることを想定して、以下の2つの方法のうち少なくとも1つにおいてその影響を軽減することが可能であろう。(1) ベースライン時の疼痛に関する特定の組み入れ基準に対して試験責任医師と患者の盲検化を維持する方法、(2) ベースライン時の疼痛とは無関係に全ての患者を無作為割り付けするが(これによってベースライン時のスコアを膨らませようとする金銭的および個人的なインセンティブが排除される)、主要有効性解析はベースライン時の疼痛が一定強度を上回る患者のみを検討することを予め定める方法。後者のアプローチを用いる場合は、ベースライン時の疼痛強度が事前に設定したカットオフ値を上回る被験者数を十分確保するために無作為割り付けに層別化の必要がある。しかし、これらのアプローチの追加的コストを分析感度に対する潜在的ベネフィットと比較検討する必要がある。

ベースライン時の疼痛強度が極めて高い患者(例 0～10のNRSで9以上)については、そのような極めて高いレベルの疼痛はおそらく心理社会的ストレスを反映したものであり、これらの患者が治療に反応する可能性は低いと考えられることから、除外を考慮することもある。さらに、例えば、ベースライン時の疼痛強度の評価が0～10のNRSで平均4以上であるが、疼痛日誌で3未満の評価が何日かあるようなベースライン時の疼痛強度の評価が極めて変化しやすい患者は除外できる。有痛性DPNに関する最近の3件の試験で、7日間のベースライン期間中の疼痛日誌の評価に25%を上回る変動がある患者を除外しているが[14,25,51]、これらのうち2件の試験では有意な治療効果が示されなかったため、このアプローチの潜在的ベネフィットは依然として不明である[14,51]。同様に、ベースライン時の疼痛強度の評価の変動が最小限または変動がなかった患者は疼痛変化を見分ける能力が限られていることもあり、これらの患者は臨床試験からの除外を考慮する必要がある。しかしながら、疼痛強度の極端な変動性(または安定性)は、根底にある病理生理学的機序の差を反映している可能性があり、そのような患者を除外すれば試験結果の一般化が限られるかもしれない。

ベースライン時の疼痛強度が鎮痛薬に関する臨床試験のアウトカムに及ぼす影響については、既存の臨床試験のデータベースをレトロスペクティブに分析することによって、例えば、ベースライン時の疼痛強度の変動の大小が分析感度の改善と関連しているかどうかを判定することによってさらに理解が深まった。この疑問をプロスペクティブに検討するために、ベースライン時の異なる疼痛強度(または変動性)を有する被験者を対象に治療効果を比較する予め定められたサブグループ解析の実施が、今後の慢性疼痛の概念実証試験や確認臨床試験で可能と思われる。

慢性腰痛およびOAに関するRCTでは「フレアデザイン」がしばしば使用されており、このデザインには、全ての鎮痛薬を中止した後に患者の疼痛レベルおよび/またはそれに関連した障害がある程度増すに違いない患者、および事前に設定した基準を満たした患者のみが無作為割り付けされている[6,71]。この戦略によって、ベースライン時の疼痛強度が高く設定されるだけでなく、非ステロイド性抗炎症薬やオピオイド系鎮痛薬などの前治療効果のあった患者および同一の作用機序を有する別

の薬剤の効果がみられる可能性のある患者を特定できる。

### 3.5. 内科的および精神科的な併存疾患

慢性疼痛に関する臨床試験では多数の内科的および精神科的な併存疾患が除外基準とされている。これらは調査対象の治療および病状によって異なり、また被験者の安全性の確保、不均一性の軽減、研究結果の解釈に際してのあいまいさの回避を目的にプロトコールにしばしば含まれる。

慢性疼痛の患者は複数のタイプの疼痛を経験することが多く、慢性疼痛に関する RCT の診断基準を満たすことが他の疼痛症状の存在を排除することにはならない。研究対象症状以外の疼痛症状が存在すれば、特に標的疼痛症状に対する治療効果の評価に際して被験者の評価能力を妨げるかもしれない。例えば、痛性 DPN 患者は腰椎神経根障害も併発している可能性があり、FM 患者は慢性腰痛がある場合が多く、併存疾患に対して異なる治療効果があると標的疼痛の報告する際に混乱が起こる。したがって慢性疼痛試験では、治療対象の疼痛よりも重度の疼痛、同一部位の疼痛または治療対象疼痛と明確に区別できない疼痛を有する患者は除外されることが多い。さらに、標的疼痛を増強させる可能性があるその他の内科的症候、または評価を妨げる可能性がある症候、例えば PHN または DPN における感覚に影響を及ぼす皮膚症状[114]や OA 試験における通風の病歴[6]は除外できる。

慢性疼痛試験に組み入れられた被験者の安全性を確保し、発現する有害事象の解釈を促すために通常、追加的な臨床的除外基準が必要とされるが、これらは研究対象となる特定治療、ならびにそのリスク、薬理学、その他の治療との相互作用に関する知見によって異なる。そのような除外基準の例として、非ステロイド性抗炎症薬に関する試験での胃腸潰瘍の存在や、主に腎排泄により消失する薬剤に関する試験での低いクレアチニン・クリアランス値および腎機能不全の既往がある。

多数の慢性疼痛試験の除外基準として、精神障害、大うつ病、不安障害、自殺念慮または自殺企図、薬物乱用またはアルコール依存、認知機能障害などの神経精神医学的併存疾患がある。これらの併存疾患の既往歴がある、または現在これらの併存疾患がある人の参加は、少なくとも以下の 3 つの重大な懸念が考えられる：(1) 精神障害または認知機能障害のある人は患者の自己報告によるアウトカム指標の評価を正確に記入できない、(2) 自殺リスクのある人や意図的または偶発的に被験薬を過量摂取するリスクのある人、薬物乱用の既往歴のある人、および乱用の可能性や突然中止すると退薬症状の発現リスクのある治療薬を服用している人、(3) 抑うつ、不安、その他の心理社会的苦痛のある患者にプラセボ反応が増強され、また／または実薬治療反応が低下することによって臨床試験の分析感度が低下する可能性。単回投与試験の結果から、心理社会的苦痛が高い被験者はプラセボ反応が高く、オピオイド系鎮痛薬反応が低いいため、苦痛が低い被験者よりも総合的な治療効果が小さくなることが示されている[149]。しかし、FM に関する最近の研究から、治療効果が抑うつ診断[120]や不安および抑うつ症状[4]によって異なることを示すエビデンスはない。しかし、抑うつや不安が分析感度に及ぼすマイナス影響があったとしても、FM 治療に用いられた薬剤がこれらの精神科的症候に及ぼす有益効果によって相殺された可能性がある[102]。

臨床試験の被験者の安全性を確保することは最優先事項であり、治療によって重篤な有害事象を引き起こすことが予想される併存疾患のある患者は除外すべきである。一般に、重症または重篤な有

害事象の発現率が高いと予想される患者を除外すると試験終了前の脱落率を低下させ、そうすることによって分析感度が高まる。しかしながら、臨床試験で除外基準数が多くなるほど、研究サンプルが抽出される地域社会の患者集団に対する得られた結果の一般化が低下し、よくみられる併存疾患があるような患者に対する治療リスク(およびベネフィット)についての学習経験が低下する。このように、安全性、分析感度、一般化の間には極めて重要なトレードオフ(相殺)が存在し、慢性疼痛に関する臨床試験へのこのような除外基準の使用を検討する際には極めて慎重に評価する必要がある[116]。

### 3.6. 併用薬およびレスキュー薬

被験者に対し安定用量の併用鎮痛薬の継続を認めるかどうか、これらの使用を除外するかどうか(特定の種類のレスキュー鎮痛薬を除く)についての問題は、慢性疼痛の確認臨床試験のデザインに重大な課題を提起している。併用鎮痛薬を被験者が継続できる臨床試験では、併用鎮痛薬が禁止されている試験に比べて、プラセボと比較した場合の追加的ベネフィットの証明の確率が低くなるとしばしば考えられている。一方、さまざまな慢性疼痛症状に適応される安全で有効な第一選択薬とみなされるような薬剤が増えつつある。これらの薬剤が広く利用できる地域では、いずれも服用していない患者やこれらの服用を簡単に中止すると思われる患者は一般に難治性であり、したがって新規の治療に効果を示す可能性が低くなると考えられる。このような患者を臨床試験に組入れると有効性が証明される可能性が低くなるかもしれない。さらに、慢性疼痛試験で他の鎮痛薬の併用を禁じれば、併用療法が一般的である日常の臨床診療の現場で得られた結果を一般化できる可能性が少なくなるかもしれない。

残念ながら、慢性疼痛試験の分析感度に及ぼす併用鎮痛薬の影響を評価したデータは限られている。4 件の PHN 試験のメタアナリシスから、鎮痛薬を併用した被験者と併用しない被験者を対象としたプレガバリンの治療効果をプラセボと比較したところ同等であることが判明したが、これらは事後解析であった[28]。高濃度のカプサイシン・パッチに関する最近の RCT から、試験期間中に 1 種類以上の鎮痛薬の服用を継続した被験者と、そのような薬剤を服用しなかった被験者においてプラセボと比較して有意なベネフィットが認められた[8]。さまざまな併用鎮痛薬が安定用量で投与されている被験者とそうでない被験者を対象とした治療効果の比較を事前設定した解析により、併用薬が慢性疼痛試験に及ぼす影響への理解が深まった[8,132]。

治療効果の検出が改善すると考えられる理由やその他の理由(例えば、忍容性および安全性の評価の解釈が容易となる)により併用鎮痛薬を除外すれば、プラセボ群の倫理的利用は通常、レスキュー鎮痛薬の使用を必要とし、これによって許容できない疼痛を経験する被験者にも試験継続が可能になる。例えば、神経障害性疼痛に関する多数の試験では、アセトアミノフェン(パラセタモール)がレスキュー鎮痛薬(最高用量は 2~4 g/日が多い)として認められている。このようなレスキュー鎮痛薬が許容される場合、被験者が使用した量を記録および報告することが重要である。有効治療に関する試験のプラセボ群ではレスキュー鎮痛薬の量は多くなる可能性があり、データの分析および解釈ではこの点を考慮すべきである。

アセトアミノフェンなどのレスキュー薬は低用量であっても、レスキュー薬の薬理効果およびプラセボ効果は実薬治療群とプラセボ群の間での被験者反応の差を弱める可能性がある。被験者がレスキ



ユー薬を必要とするような疼痛が発現した場合、すみやかに治療を中止することを奨励しているならば、倫理的な面からレスキュー薬を認めないということもあり得るかもしれない。そのような試験は、耐え難い疼痛による中止までの期間に実薬群とプラセボ群の間で差を検討するように設計できる。レスキュー薬の禁止に対してそのようなアプローチを用いた慢性疼痛試験は少ないことから、確認試験でそのような試験デザインを考慮する前に追加的研究が必要となるであろう。

併用鎮痛薬(レスキュー薬以外のもの)を禁じる場合、ベースライン期間の開始前に除外対象の薬剤を安全に中止すると共に、それらの薬剤の影響(すべての退薬症状または持続中の治療効果)の停止が確保できるよう十分な休薬期間がなければならない。これによってベースラインのデータが薬剤中止の一過性の影響を受けないようにすることができる。このような休薬のために必要とされる期間は、除外対象の薬剤の薬物動態学および薬力学的な特徴によって大きく異なる。

併用鎮痛薬を認める場合、被験薬および市販薬を含む許可された薬剤すべての薬物動態学および薬力学を徹底的に評価することにより潜在的な相互作用を明らかにしなければならない。被験薬に対する禁忌薬は全て除外すべきであり、全ての併用鎮痛薬は組入れ前の一定期間および試験期間中に安定用量を維持する必要があり、医学的に必要となる場合を除いて試験期間中に中止してはならない。さらに、試験開始後は鎮痛薬による治療を開始してはならない。これは慢性疼痛患者が用いることが多いその他の疼痛療法(例えば、理学療法、心理社会的介入、侵襲性介入)にも適用される。特定状況の下では、試験期間中における安定用量の併用鎮痛薬使用が治療反応に及ぼす影響をベースライン共変量として、または併用鎮痛薬を投与された被験者と投与されていない被験者を比較するサブグループ解析による有効性アウトカムの統計解析によって検証できる[8,58]。

慢性疼痛試験において併用鎮痛薬を認めるための異なるアプローチがある。1つのアプローチは、全ての鎮痛薬について試験期間を通して安定用量の継続を認めるというものである。2つ目は、既存治療薬の安定用量に被験薬を「上乗せ」して検討するものである。規制当局への申請に関連して実施される試験について、最もわかりやすいアプローチは、研究対象の特定の慢性疼痛症状に対して国または地域当局から承認されている既存の治療を用いることである。ただし、特定の慢性疼痛症状に対して規制当局から承認されていないが、一般に有効と考えられている薬剤(例えば、FM または神経障害性疼痛に対する三環系抗うつ薬)があるが、承認されている適応は当該規制当局によって異なる場合が多い。したがって、3つ目のアプローチは、限られた数の特定の鎮痛薬(例えば、規制当局が承認したもの、または有効性のエビデンスが一般に受け入れられているもの)のみを、試験期間を通して安定用量で使用することを認め、それぞれについて最高用量を明記するというものである。例えば、有痛性DPN患者を対象とした試験では、以下の薬剤のうち2つに限って安定用量で使用を認めることができるかもしれない:(1) ガバペンチンまたはプレガバリン、(2) オピオイド系鎮痛薬またはトラマドール、または(3) 三環系抗うつ薬または選択的セロトニン・ノルエピネフリン再取り込み阻害薬。もちろん、使用はこれらの3種類の薬剤のうち1種類のみに限定するか3種類すべてとし、有効用量を最小限、適度または最大限に反映できる最高用量を事前に設定することができる[11]。しかし、併用鎮痛薬の許可に関してこのような構造化されたアプローチを用いている臨床試験はわずかであることから、このような意思決定の基となっているデータは少ない[132]。

慢性疼痛に関する確認試験の設計では、レスキュー薬の持続期間が限られており、理想的には主

要有効性アウトカムと一致しないことを条件として、被験者に無関係な急性疼痛症状(例えば、歯科治療や外傷)が発現した場合にレスキュー薬の許可を考慮することができる。

### 3.7. 前治療に対する反応

現在、作用機序が異なる多種多様な薬剤がさまざまな慢性疼痛症状に適応されており、有効性に関するエビデンスが広く受け入れられているが、十分でない鎮痛効果、忍容性不良または不十分な用量漸増によって地域社会で相当数の患者で十分な治療効果がみられていない[53]。このような患者が難治性の疼痛を有している場合や一般に薬剤の副作用に敏感である場合、これらの患者を臨床試験に組入れると分析感度が低下し、有害事象例や中止例の数が増す可能性がある。

これらの理由により、新規の作用機序を有する薬剤がこのような患者に有効となる可能性を知ることが重要であるが、有効性欠如または忍容性不良により複数の前治療に効果がなかった患者の除外を考慮できる。レトロスペクティブな患者報告に基づいて有効性欠如および忍容性不良を明確にすることは可能であるが[136]、このような判定を行うことには問題が多く、薬剤用量や治療期間が十分に思いつけないことから限られる。いくつかの臨床試験の解析は、研究開始前の種々の鎮痛薬使用歴があるサブグループで治療効果を比較することにより前治療の役割を検討している[例 71]。慢性疼痛に関する臨床試験で前治療における治療反応が分析感度に及ぼす影響への知見をさらに認識するため、被験者の治療歴に関して十分な情報が収集されている終了している試験を対象としたレトロスペクティブな評価を実施すべきである。

### 4. 慢性疼痛臨床試験の相

慢性疼痛の臨床試験のために患者をスクリーニングし、十分な説明に基づき患者からの同意を得た後に疼痛などの変数が評価されるのが一般的であり、続いて被験者を選択し、種々の治療群に無作為割り付けする。RCT では無作為割り付け前の期間が用いられることが多く、これらの期間は被験者への治療(無治療、プラセボ、被験薬)、盲検化の程度(非盲検、被験者を対象とした単盲検、被験者と試験担当者を対象とした二重盲検)および導入期間の目的によって異なる[101]。上述したように、このような期間は被験者が無治療または併用鎮痛薬の安定用量を投与されている期間の疼痛の重症度およびばらつきを確認するために用いられる。無作為割り付け前の期間は、被験者がプラセボで鎮痛を経験するかどうか、実薬で鎮痛や耐え難い副作用を経験するかどうかを評価するためにも用いられる。さらに、無作為割り付け前の期間は研究手順に従わない被験者を特定した後に除外するためにも利用できる[83]。例えば、最低日数分(例 7日中4日または5日)の疼痛日誌の記入を完了しない被験者は、このような早期段階のノンアドヒアランスは無作為割り付け後のその後の研究手順に対するノンアドヒアランスを予測するという想定に基づいて、無作為割り付けから除外できる[7,118]。

#### 4.1. ベースライン期間

最も簡単なタイプのベースライン期間において、被験者にはプラセボも実薬も投与されず、無作為割り付けへの適格性を確認し、その後の治療効果を評価するためのベースラインとするためにデータ

が収集される。このようなベースライン期間中には、主要および副次的有効性アウトカム指標の他に、試験期間中に発現する有害事象がベースライン期間にすでに存在していたか「治療下で発現した」かどうかを確認するための症状、バイタルサイン、臨床検査値、その他の臨床評価をはじめとする種々の評価が収集可能である。十分な数の被験者の無作為割り付けは、大半の場合、ベースライン時の疼痛、その他の重要な人口統計学的特徴および臨床的特徴に群間での差を制限するが、不均衡は偶発することもあり、データの分析および解釈に際してはこのことを考慮すべきである[130,147]。

慢性疼痛試験では通常 1 週間のベースライン期間が用いられている。しかし、疼痛などの変数のベースライン評価の信頼性は、ベースライン期間が長くなって測定数が多くなると大きくなる可能性がある[64]。ベースライン期間が長くなると無作為割り付け前に疼痛強度が低下する機会が増えることにもなる。このような低下は平均値への回帰によって起こることがあるが[150]、この原因として考えられるのは、測定の信頼性が欠如しているか、あるいは患者の状態が悪い時に臨床試験に組入れられ無作為割り付け直後に疼痛が特徴的なレベルにまで戻る傾向がある。疼痛強度が高いか安定している場合に治療効果の検出が容易であるならば、ベースライン期間の疼痛強度にそのような低下がある患者を除外することで分析感度は高まると考えられる。ベースライン期間が長くなると、無作為割り付け前に生じる早期脱落例のための時間が長くなり、また被験者にプロトコールを確実に理解させ、馴染みのない研究手順（例えば、電子日誌、音声自動応答システム）を順守させるための期間を長くさせることも可能となるであろう。

しかし、ベースライン期間が長いことによる潜在的ベネフィットは、ベースライン期間中にレスキュー薬が提供されない場合に中止例が増え、被験者募集や長期試験への患者確保の難しさが増し、コストが増すことによって相殺される可能性がある。さらに、後述するように、長期試験はプラセボ群の改善が大きくなる可能性があり、これが分析感度を低下させるのかもしれない[31,107]。連日の疼痛記録を伴う 1 週間のベースライン期間は治療有効性の証明に成功した多くの試験で利用されており、慢性疼痛試験に広く推奨できる。しかし、特定条件の下では、ベースライン期間が長いことによる潜在的ベネフィットは 2 週間を考慮することにより十分正当化するに値すると思われる。

典型的に、二重盲検のフェーズ中にレスキュー鎮痛薬が提供されない慢性疼痛の RCT では、そのベースライン期間中にもレスキュー鎮痛薬は提供されない。このような診療については十分な配慮はされておらず、ベースライン期間中に全ての被験者に対してレスキュー鎮痛薬の使用が認められている（または試験の二重盲検のフェーズ中に全ての被験者に使用が認められたレスキュー鎮痛薬の用量が少ない）場合、プラセボ群では無作為割り付け後にベースライン時の疼痛レベルが急速に低下する可能性は低い。

#### 4.2. プラセボ導入期間

抑うつへの介入に関する臨床試験では単盲検導入期間が用いられることが多い。これは全被験者に対しプラセボが投与されていることを研究担当者は知っているが、被験者はそれを知らないというものである。大半の場合、（事前設定した閾値未満への低減、または事前設定した低減割合の達成またはその両方によって）症状が改善する被験者は「プラセボ反応例」とみなされて、二重盲検治療の相からは除外される。この研究デザインに関する理論的根拠は、導入期間にプラセボに反応する被

験者を除外すると無作為割り付け後のプラセボ群で反応率が低下し、それによって治療効果検出に関する試験の感度が高まるとの仮定に基づいている[27]。慢性疼痛試験のプラセボ群では大きな反応があることが多いため[31,68,107]、このような単盲検導入期間が検討可能である。しかし、抑うつに関する試験で単盲検プラセボが導入期間を採用したものとしなものを比較したメタアナリシスの結果、この戦略がプラセボ群の反応を有意に低下させた証拠も分析感度を有意に高めた証拠も認めなかった[88,91,109,142]。導入期間中に改善する被験者が、プラセボへの無作為割り付け後に同様の反応を示すエビデンスはほとんどなく、このような導入期間が分析感度を上げることに失敗する理由がこのことによって説明できるようである。

おそらく精神医学領域での試験で残念な経験があることから慢性疼痛試験ではプラセボ導入期間が使用されることは稀であり、したがってこの種の導入期間が分析感度を高めるかどうかは不明である。DPN に関する最近の試験で、症状強度および変動性の基準を満たさなかった患者を除外するための 1 週間の単盲検プラセボが導入期間を設けたにもかかわらず、プラセボ群の被験者の 26%が 50%以上の症状軽減を経験したが[156]、プラセボ群の反応はこのような導入期間がなければさらに高かった可能性がある。腰部神経根障害に関する最近の試験で、無作為割り付け前に単盲検プラセボ導入相と単盲検実薬治療相の両方を用いてプラセボに対する実薬の優越性を証明することに失敗したが[103]、これらの否定的な結果の説明として、分析感度を損なうようなその他の研究因子またはこの疾患における薬剤の有効性欠如が考えられる。

単盲検プラセボ導入相に代わるものが二重盲検可変期間プラセボ導入相で、この期間は治療開始に関連する有効性（および忍容性）の評価におけるバイアス低減を目的として、試験責任医師と被験者の両者が実薬治療の開始について盲検化される[40]。治療終了時に二重盲検終了相が含まれる場合もある。いくつかの最近の研究は、無作為割り付け前にプラセボ反応例を特定し、これらを試験または主要有効性解析から除外する[40,49]と共に、患者および試験責任医師を実薬治療の開始および終了について盲検化する[23,24]ために可変期間導入相を利用している。このような導入期間が抑うつ薬による治療効果の検出力を高める可能性があるとする多少のエビデンスがあるが、二重盲検可変期間プラセボ導入相を慢性疼痛試験に含める価値があるかどうかは前例がないことから明らかでない。したがって、このデザインの特徴を確認試験に含める前に初期相の試験で検証することは無駄ではないと思われる。

重要なことは、プラセボ導入期間には試験の治療効果検出能を高める可能性以外の目的が考えられるという点である。例えば、FM に関する試験では単盲検プラセボ導入期間が「安全性指標を含むベースライン変数に関するバイアスのない評価を得るために」用いられている[5]。

#### 4.3. 無作為割り付け前の強化 (enrichment)

RCT における無作為割り付け前に被験者サンプルを強化するには複数のアプローチがあり、すでに述べたように、これには以下の期間が含まれる：(1) アドヒアランスが不良となる可能性が高い被験者を特定するためのベースライン期間、(2) プラセボ反応例である可能性が高い被験者を特定するプラセボ導入期間。被験者サンプルを強化するためのもう一つのアプローチは、まず全被験者に実薬治療を行い、これはほとんどの場合、オープンラベルまたは単盲検法で実施される[89]。これによって、

治療に反応した被験者は、次の二重盲検相で反応する可能性が高く、したがって分析感度を高めるとみなされた被験者のみを無作為割り付けできる(セクション 6.3 参照)。無作為割り付け前に被験者を実薬治療に曝露させることで、治療に忍容性のある被験者のみを無作為割り付けすることも可能となるが、これが二重盲検相の有害事象および中止例の発現を低減させ、それによって欠測値の数が抑えられると考えられる。

実薬治療またはそれと強い関連性のある治療が無効であった患者についても無作為割り付けから除外可能である[例 114]。しかし、ガバペンチンおよびプレガバリンに関する試験のメタアナリシスで、強化の度合いは最大で無作為割り付けした被験者の 12%と推定されたが、これらの薬剤のうち 1 種類が無効であった臨床歴に基づく患者の除外(全患者が治療に曝露されるわけではないため「部分」強化の一種)は、強化無しの場合と比較して、有効性や有害事象の推定値を変化させることを示すエビデンスは認められなかった[137]。

さまざまな種類の強化戦略は、実薬に対する反応の可能性を高めたり、プラセボ群の反応を低下させたり、有害事象によって試験を中止する被験者数を抑えたりすることによって分析感度を改善すると思われる。また、おそらく明確な疼痛機序もしくは薬物動態または薬力学における個人差のために、反応例からなるサブグループがある場合、強化は特に有用かもしれない[17]。しかし、治療効果の推定値は治療に反応した患者からしか得られず[89,121]、被験者が無作為割り付け前に実薬治療による副作用に気付いたために盲検化が損なわれた場合にはバイアスが生じる可能性がある[87]。このことから、強化は臨床試験結果の一般化の可能性を弱める。しかしながら、強化を用いた臨床試験デザインは後述するように、慢性疼痛試験で次第に注目を集めている[69,70,96,106]。

#### 4.4. 漸増相、維持相、漸減相

慢性疼痛に関して検討されている薬剤の多くは、有効用量を達成するために漸増相を設けて副作用の発現率および重症度ならびに特定の重篤有害事象(例 Stevens-Johnson 症候群)の発現を最小限に抑えられるようにすることが要求されている。この漸増相の期間は薬物動態および標的用量によって異なるが、1 週間から 8~10 週間の範囲であった。他の条件がすべて同等であれば、鎮痛が迅速に達成されるためだけでなく、後述の通り試験期間が短いほど分析感度が高くなる可能性があるために漸増相が短いことは利点である。

慢性疼痛の確認試験における漸増相の完了に続く治療相、すなわち維持相の持続期間は過去 20 年間に長期化している。初期の慢性疼痛試験では、維持相の設定が 4~8 週間の場合が多かったが、12 週間の維持相は規制当局の要件を満たすことを目的とした確認試験の標準となった。症状が慢性的であることを考えると、患者はこれらの治療薬を数ヵ月または場合によっては数年にわたって用いることになるため、慢性症状に対する治療は相当期間における持続的な有効性および安全性を証明しなければならぬと言える。このことは、有効性および安全性に関する短期間の評価は、長期間の評価ほど正確に長期治療効果を予測しないと仮定したものである。しかし、その妥当性に関する系統的な試験の欠如、種々の鎮痛薬による晩期副作用の発現、ならびに慢性疼痛に対する特定治療に対して患者に耐性になることという考えを前提にすると、この仮定は妥当である。試験期間が長いほど慢性疼痛治療薬の長期効果をより良く評価できることが示唆されるが、不十分な鎮痛を原因と

するプラセボ群からの脱落例数は試験継続期間を増加させると考えられるため、期間延長を伴ったプラセボ対照試験は問題となる。ただし、特定の研究目的について、例えば、ある薬剤の鎮痛作用への耐性を検討するためには長い試験期間は必須と思われる。

大うつ病[76]および DPN の疾患進行[139]に関する臨床試験では、プラセボ群の反応の改善は持続可能で、経時的に増強することを示すエビデンスがある。重要なことは、神経障害性疼痛試験のプラセボ群における反応は、長期と短期を比較した試験でも同じ様に大きいと考えられることが最近明らかになっている。これはおそらく「19 週までの時点においてもプラセボへの反応がプラトーに達していることを示す徴候がなかった」[107]ことによる。プラセボ群の反応のこのような増強の理由は不明であるが、試験担当者との接触および試験参加によるその他の非特異的な影響がプラセボ群における持続的な改善をもたらした可能性はある。プラセボ群における改善が継続し、その後実薬治療による改善に近づく一方で、疼痛がめったに軽減されない「床」より下のレベルに有効な治療がより急速に到達するならば、慢性疼痛に関する長期試験では分析感度が低下している可能性がある。これらの考察および現行の規制当局のアプローチに基づき、慢性疼痛治療に関する確認試験については 12 週間の維持相が推奨されるが、特定状況の下では、これより短い期間または長い期間が必要とされる場合がある。

慢性疼痛に関する RCT の維持相が完了したら、被験者は二重盲検治療および投与されていた場合はレスキュー鎮痛薬についても突然中止するか漸減治療を受けることができる。突然の治療中止が退薬症状や中断症候群を引き起こす可能性がある場合は試験に漸減相を含めることが必要であるが、漸減相はそれらを予防または最小限に抑えるのに十分な長さとするべきである。漸減相は、治療を中止した時に実薬治療を受けていた被験者における治療ベネフィットが、プラセボ投与と被験者に比べて大きく減少するかどうかを検証する機会にもなる。しかし、特定状況の下では、漸減相を設けずに実薬治療を終了すると退薬症状が発現するかどうかを対照を置いて検証する機会となるが、これは薬物療法の即時中止を必要とするような有害事象が発現する場合には臨床的に重要な問題である。治療中止後も有益作用が持続する可能性があると考えられる場合、維持相に続く期間は実薬群の被験者をプラセボに切替えてその可能性を評価する期間とみなすことができる。

#### 5. 治療群

慢性疼痛の確認試験では、被験薬による治療患者群を 1 群以上置くことに加えて、ほとんどの場合でプラセボ群を置き、場合によっては実薬対照群も含めることがある[37,38]。実薬対照群は一部の地域では規制当局の要件を満たすために必要とされる場合がある。初期相の試験の焦点であることの方が多く、非有効量、最小有効量または最大耐用量を確認する必要から、種々の用量の実薬を 1 群以上の追加群に投与する必要があるかもしれない。しかし、抑うつ試験解析では、治療群が少ない試験では可変用量戦略を用いる傾向も強かったが、多数の治療群を置くことがプラセボとの識別が上手くいかなかったことに関連していた点は注目に値する[75]。患者は改善を期待しており、したがって患者がプラセボを投与されている可能性が低いことを知っている場合にはプラセボ群の反応が大きくなることから、多数の実薬治療群を置いた試験は分析感度が低い可能性がある[123]。さらに、1 件の試験で複数の治療群を含める場合は、複数の比較に対応する計画を予め設定しておく必要があり、

これにはサンプルサイズを大きくする必要が生じることがある[98,145]。一般に、慢性疼痛に関する確認試験では治療群の数を限定し、出来る限り最大でも3群、すなわち、実薬治療群、プラセボ群および第2の用量群または実薬対照群までとすることが推奨されている。

### 5.1. プラセボ群

治療開始の遅延が持続性の副作用の原因とならない場合、また、被験者がどの時点でも理由を問わず試験を中止できる権利を有していることを十分に理解しているという仮定の下で、被験薬とプラセボの比較は有効性および安全性を評価する上でゴールドスタンダードとみなされている[35,60,86,138]。プラセボ群は同時に内部標準として被験薬の有効性および安全性の評価を可能にし、この内部標準は、有効性が確立された治療がプラセボに対する優越性を常に示さず[42,69]、症状の重症度さらには自然消失において著しい変動が起こり得る慢性疼痛などの症状を検討する際に必要とされる。

慢性疼痛試験におけるプラセボ群の改善の程度はさまざまで、場合によっては相当な程度になることがあり、例えば、神経障害性疼痛に関する試験では、ベースライン時から試験終了時までには50%以上の疼痛軽減を経験する被験者のうち0%から約60%までと幅がある[42,68]。疼痛試験のプラセボ群の反応は、プラセボ効果だけでなく自然歴や自然消失ならびに上述の平均値への回帰を反映しているという点を十分に理解することが重要である。プラセボ群の反応の大きさと、試験がマイナスのアウトカムを有している可能性が高いことの関連性は、抑うつおよび神経障害性疼痛[27,31,68]、その他の内科的および精神科的疾患の治療に関する臨床試験で認められている。しかし、この関連性は平均値への回帰および統計解析において適切に説明されない測定誤差の影響を反映している可能性がある[3,128]。

プラセボ群の反応と試験アウトカムの関連から、抑うつ[27]および慢性疼痛[31,39,61,68,107,155]に関する臨床試験におけるプラセボ群のより大きな反応と関連性のある因子の特定が検討されてきた。このような取り組みの主要な目的の1つは、プラセボ群の反応の軽減が可能となり、それによって、真の治療効果が存在する場合にプラセボとの比較において実薬の優越性が試験によって示される可能性が高まるような修正可能な試験の特徴を特定することである。しかし、プラセボ群の反応を軽減するような因子はいずれも実薬への反応の低下とも関連していると思われる、そのためプラセボ群の反応が軽減することは必ずしも、試験が治療ベネフィットの証明能力を高めることにならないと考えられる。

多数の鎮痛薬にみられる副作用(例 便秘、鎮静、悪心)は、被験者と試験責任医師ともに割り付けられた群の患者にみられる副作用を推測する可能性を高めると言える。インフォームドコンセントに記載されている薬剤による副作用を経験しているプラセボ群被験者の「ノセボ」効果はこのような盲検解除を相殺すると考えられる[99]。しかしながら、被験者が実薬を投与されていることをうまく推測した場合の症状改善に対する期待から、副作用による盲検解除はRCTの内部妥当性を損なう可能性があり、治療を実際よりも有効にみせることになるかもしれない。この可能性に対処するには少なくとも2つの方法論的アプローチがある。1つは、試験終了時に被験者が割り付けられていた治療群についてどのように推測しているか、その推測に至った主な理由について被験者と試験責任医師に尋ねる

評価を実施して[29,146]、試験期間中に有意な盲検解除があったかどうかを判定し、もしあったならば、それが実薬の有効性、副作用またはその他の原因によるものかどうかを判定する。実薬に関連した鎮痛に基づいて治療群に関する正確な推測が高率で行われることは問題とはならないが、副作用、その他の因子に基づく正確な推測は治療効果の推定にバイアスをもたらす可能性がある。このような評価は慢性疼痛に関する確認試験では通常行われてこなかったため、盲検解除の証拠がある試験結果の解釈に関する指針はない。しかしながら、試験終了時点で被験者が割り付けられていた治療群に関する被験者および試験責任医師の推測ならびにその推測に至った主な理由の評価を検討することが推奨される。

副作用からの盲検解除の可能性に対処するもう1つのアプローチは、通常は不活性なプラセボを用いる代わりに、実薬の副作用を模倣しているが鎮痛効果がない「アクティブプラセボ」を利用するというものである。鎮静に関連した薬剤に関する臨床試験では、ベンズトロピン、ジフェンヒドรามין、ロラゼパムなどの薬剤の低用量投与がこの目的のために用いられている[例 46]。同様に、オピオイド系鎮痛薬に関するRCTでは、低用量の塩酸ジフェノキシラート/硫酸アトロピンなどの便秘を引き起こす薬剤がアクティブプラセボとして利用可能である。アクティブプラセボの利用は副作用による盲検解除を減少させるが、アクティブプラセボの影響により有害事象や中止の発現率は上昇するかもしれない。また、ベネフィットがないが潜在的悪影響がある薬剤に被験者を曝露するという倫理的側面については慎重な検討が必要であろう。これらの理由から、アクティブプラセボの利用は慢性疼痛に関する確認試験では推奨されないことが通例である。

慢性疼痛を含むさまざまな疾患の治療に関するRCTでのプラセボ群の利用について極めて切実な意見があるにもかかわらず[35,138]、有効な治療が存在する疾患に関するRCTにプラセボ群を用いることの倫理上の問題に関する議論は続いている。現在のところ、この問題に関する主要な倫理委員会や規制当局による勧告はすべて、そのような設定下でのプラセボ群は、被験者が十分な情報を与えられ、研究期間中のプラセボによる治療が被験者の長期的健康に影響を及ぼすことが予想されない場合に限って許容される、という立場をとっている。

### 5.2. 実薬対照

プラセボに対する被験薬の優越性の証明に失敗した場合、その原因が有効性の欠如(すなわち「薬剤の失敗」)にあるのか有効性の証明能力を損ねたRCTの方法論的特徴(すなわち「試験の失敗」)にあるかが不明確である。慢性疼痛試験ではこのような状況が頻繁に発生するようになっており、実薬対照、すなわち、プラセボに加えて有効性が十分に確立された治療薬を含めることで、実薬対照がプラセボに比べて有効であることが示されれば、当該試験が十分な分析感度を備えていたと結論付けることが可能となる。

さらに、一部の地域の規制当局は、治療に関して確立された見解がある場合には侵害受容性疼痛および神経障害性疼痛に関するRCTに実薬対照を含めることを推奨している[37,38]。慢性疼痛に関する確認試験に実薬対照を含めることによって、有効性、安全性、忍容性および総合的な費用対効果の比較評価の基準となるデータが得られる。上述の検討から、適切な対照が存在し、リソース面から許され、そのことによって試験の実行可能性や分析感度に影響を及ぼすレベルまで治療群の数を

増やすことがない場合は、慢性疼痛に関する確認試験に実薬対照を含めることが推奨できる(上記参照)。

実薬対照は、現行の標準ケアとみなされている有効治療に対し被験薬が非劣性、または同等のいずれかであるという仮説の検証を可能にする[50,55,59,60,73,105]。非劣性または同等性を証明するには、被験薬と対照の差は、臨床的検討および先行試験において対照によって証明された治療効果に基づくことが多い予め定めたマージンの範囲内におさまらなければならない。ほとんどの場合、同等性または非劣性の検証では優越性の検証よりも大きなサンプルサイズが必要とされる。しかし、慢性疼痛に有効な薬剤のプラセボに対する優越性は極めてばらつきが大きく、マイナスのアウトカムは稀であるため、有効な対照が実際にどのような試験においても有益な効果をもたらすかについては保証がない。そのため、被験薬と対照薬のいずれもプラセボより優れていない場合に、対照薬に対する被験薬の非劣性または同等性を証明することができる [85,138]。これらの理由から、試験にプラセボ群が含まれていない場合の非劣性または同等性の検証では慢性疼痛治療薬の有効性を十分に証明できず、プラセボ不在下では実薬対照との比較における優越性の検証が必要となる。

### 5.3. 固定用量群 vs 可変用量群

固定用量群は、ほとんどの場合、初期相の試験で確認される非有効量、最小有効量および最大耐用量の立証に必要であると一般に考えられている[57]。しかし、治療濃度域は、年齢、性別、遺伝子、体重ならびに薬剤と疾患間の相互作用におけるばらつきによって生じる鎮痛薬の吸収、代謝、生理学的分布におけるばらつきにより、同一薬剤に曝露している患者間でかなり異なる可能性がある。そのため、効果および忍容性にとって可変用量は一部の人には、治療濃度域を超える可能性があり、有害事象による中止を引き起こしたり、他の人の場合には治療濃度域に達しなかったり有効性の欠如による中止を引き起こしたりする固定用量に比べて有利となるかもしれない。さらに、被験者に投与される用量が有効性と忍容性の両方に基づいて調整されることから可変用量は臨床診療を反映したものととなっている。

抗うつ薬に関する臨床試験の評価から、可変用量デザインは固定用量デザインに比べてプラセボと比較した場合の治療の差が大きく、これは主に可変用量試験におけるプラセボ群の反応が低減した結果であることが判明している[74,75]。慢性疼痛に対するオピオイド系薬に関する試験の評価から、個別設定または調節可能な投与方法による併用漸増法は固定用量試験よりも強度に陽性となる可能性が高いことが判明した[69]。固定用量群と可変用量群を含んだ神経障害性疼痛に関する最近の試験では、プラセボとの比較において同様の改善を示し、可変用量群では有害事象による中止例が少なかった[44,135]。総合して考えると、これらのデータは、可変用量は固定用量よりも分析感度が高い他に臨床診療への結果の一般化がしやすい可能性があることを示唆している(但し、プレガバリンの可変用量を用いた最近の2件の確認試験では、研究対象とした特定の疾患における有効性の欠如を反映している可能性があるものの陰性であった[103,133]点に留意すべきである)。これらの理由から、多くの場合において固定用量に基づいた有効性および安全性のデータを補完する貴重な情報を提供するものとして可変用量の評価が推奨できる。

薬剤を可変的に投与するにはさまざまなアプローチが利用可能であり、オピオイド系薬に関する試

験のメタアナリシスでは固定-個別化、固定-調節可能および可変投与デザインが区別されており[69]、慢性疼痛試験にも検討可能である。特定の投与戦略の選択は、当該薬剤の薬物動態学的および薬力学的な既知の特徴ならびに治療濃度域において予想される個人間変動の程度に基づいて行われるべきである。可変用量に用いられる具体的なアルゴリズムは研究プロトコールの中で予め明確に設定しておくだけでなく、試験結果の公表時にも明示すべきである。

### 6. 確認試験の種類

無作為割り付けと盲検化は確認試験に用いられる試験デザインの重要な特徴であり、臨床試験の質に関する最も重要な決定要因である[63]。研究プロトコールでは、試験群への被験者の割り付け(例 2種類用量の実薬投与群およびプラセボ投与群に1:1:1の比で割り付け、実薬投与群とプラセボ投与群に2:1の比で割り付け)、無作為割り付けを遂行するために用いた具体的なアプローチ、無作為割り付け計画の特別な特徴(例 層別化、ブロック化、動的割り付け)、被験者と研究担当者の全員の十分な盲検化を確保するために用いた手順および研究期間中に盲検解除が認められている場合にはその条件および方法を予め設定し、試験報告書ではそれらを明確に記述しなければならない。被験者と研究担当者を治療に関して盲検化することに加えて、選択バイアスのリスクをさらに低減するために、治療順の割り付けもこれらの研究担当者が治療の割り付けについて予測できないように行われなければならない[19]。

### 6.1. 並行群間比較試験

慢性疼痛治療に関する確認試験で最も頻繁に利用されているのは並行群間比較デザインである。これらの試験では、患者を1種類以上の用量の被験薬投与群またはプラセボ群に無作為割り付けした後に、有効性、安全性および忍容性を評価する(図1)。有効性は、実薬による鎮痛がプラセボによる場合に比べて有意に優れていることによって証明される。上述の通り、一部の試験では実薬対照群への無作為割り付けを含み、並行群間比較デザインでは、被験者に対し安定用量の実薬投与が継続されることが多い維持相の前後に漸増相および漸減相を含む場合がある。

図1に示した並行群間比較デザインは慢性疼痛の確認試験の実施にあたって広く利用されているアプローチであるが、このデザインが最適とはならない可能性がある

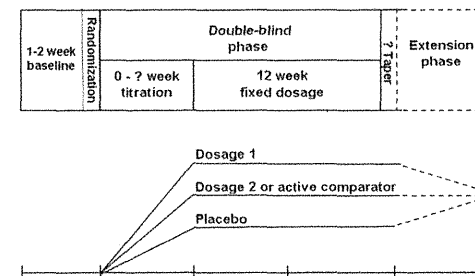


Fig. 1. Parallel groups research design for confirmatory chronic pain clinical trials.

ある慢性疼痛症状および治療がある。標準的な並行群間比較デザインが重大な限界を有する疾患には、希少疾患(例 肢端紅痛症、三叉神経痛)または12週間にわたるプラセボの継続が不可能な疾患(例 重度の癌性疼痛)がある。本セクションの残りの部分では、慢性疼痛に関する確認試験の

設計において検討可能な代替デザインについて簡潔に考察する。残念ながら、これらの代替デザインの長所および短所についてはほとんど合意が得られていない。その理由の大部分は、これらが慢性疼痛に関する確認試験に用いられた頻度は比較的低く、これらを評価するためのデータが比較的限定されていることにある。

## 6.2. クロスオーバー比較試験

クロスオーバーデザインでは、被験者を二重盲検法により 2 種類以上からなる一連の治療に無作為割り付けし、各治療の間には前治療の影響を消散させるための期間が設けられていることが多い[66,93,129]。この試験デザインでは、被験者には実薬とプラセボの両方（および場合によっては実薬対照）が投与されて自身が対照としての役割を果たすため、ばらつきが最小限に抑えられ、並行群間比較デザインの場合と比べてある治療効果を証明するのに必要な被験者数がかなり少なくなる。さらに、各患者に実薬が投与されるクロスオーバー試験では、プラセボのみを投与される可能性があることを患者が知っている並行群間比較試験よりも患者の募集が容易となる可能性がある（但し、多くのクロスオーバー試験では試験期間が長くなることによってこのメリットは相殺されるかもしれない）。このデザインはその効率の良さから初期の「概念実証」有効性試験でよく用いられているが、クロスオーバーデザインが確認試験に検討可能な状況がある。これらには、稀な慢性疼痛症状に関する研究があり、これらの症状では大規模なサンプルの募集は不可能であり、クロスオーバーデザインが効率的かつ実用的な選択肢となり得る。

しかし、クロスオーバーデザインの利用には以下をはじめとするいくつかの前提が必要である：（1）被験者は治療期間を通して安定状態にある（すなわち、進行期や寛解期ではない）疾患および疼痛症状を有しており、治療効果の推定が被験者の状態の変化による影響を受けない状態にあることが必要；および（2）1 つの治療期間から次の治療期間に移る際の持ち越し効果がないこと。クロスオーバー試験ではアウトカム指標が治療中止の直後にベースライン値に戻ることが理想的であることから、このような効果があると持続性または疾患修飾性の作用を有する治療の調査にこのデザインを用いることが不可能となる。さらに、各参加者には 2 種類以上の治療薬を投与し、2 回目の治療期間開始前に試験を中止した参加者からは治療効果に関する被験者内評価のためのデータが得られないことから、被験者の保持は非常に重要である。但し、3 種類以上の治療薬を用いるクロスオーバー試験では、2 期間以上の治療に参加する被験者からは 1 種類の治療比較に対するデータだけでも得られることになる。漸増相と漸減相を必要とする治療については、二重盲検状態を維持するために各治療期間において漸増相と漸減相を反復しなければならないため、クロスオーバー試験の長さが抑制的に働く可能性がある。漸増や漸減を必要としない治療であっても、維持相が 12 週間でなければならぬ場合、2 期間で構成されるクロスオーバー試験は最低でも 6 ヶ月間を必要とする。このような理由から、クロスオーバーデザインは慢性疼痛に関する確認試験よりも初期の試験相に適したデザインであることが多い。

## 6.3. 無作為化治療中止試験

有効性と忍容性の間に至適バランスをもたらす用量に著しいばらつきがある薬剤（例オピオイド系

鎮痛薬）を並行群間比較試験で研究する場合、副作用やベネフィットの欠如により中止する被験者は多数に上ることが考えられるが、慎重な漸増および可変投与によってこれらの中止例は低減可能となるということは頻繁に観察されてきた。無作為化治療中止デザインでは、治療薬投与を 1 回以上継続する群と中止してプラセボに替える群のいずれかに被験者を無作為割り付けする。このようなデザインは、実薬で鎮痛反応を示す被験者と副作用を許容可能でその後試験の中止相に無作為割り付けされる被験者の両方またはいずれかを特定するためにオープンラベル相または単盲検無作為割り付け前強化相と組み合わせることが多い[69,70,96,106]。主要有効性エンドポイントとしては、不十分な鎮痛による試験からの脱落までの期間[124]、無作為割り付けから試験終了時までの疼痛強度の変化[52,72]または不十分な鎮痛と「不十分な治療」に関するその他の指標の複合が考えられる[81]。試験が脱落までの期間を主要エンドポイントとして設計されている場合、このデザインは標準ケアまたは代替治療への逃げ道を組み込むことが可能であるため、プラセボ群が望ましいが長期間のプラセボへの曝露が倫理的または実的な懸念を提起する状況下では中止デザインも利用可能である。

このデザインは確認臨床試験に許容される方法とみなされており[60]、さまざまな慢性疼痛症状を有する患者を対象としたオピオイド系鎮痛薬の有効性の証明に成功している[52,69,70,72,81,124]。他の薬物または治療に関する臨床試験で無作為化治療中止デザインを用いた例は少ないが、注目すべき最近の例外としては、FM[20]、腰部神経根障害[103]および混合型末梢神経障害性疼痛[48,104]を有する患者を対象としたプレガバリンに関する RCT がある。FM 試験では、6 週間オープンラベル相が含まれ、治療に忍容性を示したプレガバリン反応例を特定し、プラセボを投与した被験者をプレガバリンを投与した被験者と比較した無作為化治療中止相では「治療反応の喪失」までの時間がより速く発現することを証明した。混合型末梢神経障害性疼痛に関する試験には 4 週間単盲検可変用量漸増相が含まれ、プラセボ群に無作為割り付けされた患者では疼痛強度が有意に高いが、中止または「疼痛の有意な増大」までの時間の群間差は非有意であることを証明した。腰部神経根障害に関する試験ではプラセボ反応例を特定するための 1 週間単盲検プラセボ導入相と 4 週間単盲検プレガバリン強化相が含まれたが、「疼痛の有意な増大」または中止までの時間におけるプラセボ群とプレガバリン群に無作為割り付けされた被験者間の差は証明できなかった。強化された無作為化治療中止デザインにおけるこのような否定的な結果は、プラセボ群に無作為割り付けされた被験者における鎮痛持続を反映している可能性があり、これは実薬またはプラセボ効果の持続的な作用から生じているのかもしれない。したがって、特定の状況下で、プラセボを投与された被験者に疼痛の有意な増大が発現するために十分な時間をとるには長期間にわたる二重盲検相が必要かもしれない。

有効かつ忍容できる用量を確認するための漸増は、無作為化治療中止デザインに強化相が含まれている場合に用いられることが多い。理想的には、プラセボに無作為割り付けされた被験者における強化相から二重盲検相への治療ベネフィットの持続は限定されたものとなる。オピオイド系鎮痛薬およびその他の治療の一部については、プラセボに無作為割り付けされた被験者の退薬症状は試験の早期中止の可能性を高めたり、もしくは実薬継続に無作為割り付けされた被験者に比べて強い疼痛の一因となることがあるかもしれない。上述の通り、強化された無作為化治療中止デザインでは、薬剤による副作用の持続またはその後の副作用の消失により被験者が治療の割り付けを正確に推測できれば無作為割り付け後に盲検解除される可能性がある[87]。上述の混合型末梢神経障害性

疼痛に関する RCT では、無作為割り付け後の盲検解除に関するこのようなリスクは、無作為化治療中止相の第 1 週目中にプラセボ群に低用量のプレガバリンを投与することによって低減した[48]。

強化された無作為化治療中止デザインは、治療によって既に改善した被験者において有効性が証明されるという明らかな「循環性」が批判される可能性があるが、Katz[69]が指摘しているように、これはデザインに関する誤解を反映したものであると考えられ、このデザインの強化相において示された治療ベネフィットは「示唆されるが、その後のプラセボ対照中止相におけるプラセボとの比較無しには確認できない」。また、総合的な治療反応率および初期の強化相における忍容できない副作用による試験中止は、臨床診療での治療方法に関する有益な特徴を示す重要な副次アウトカムである点を重視することも重要である[70]。

#### 6.4. 用量反応試験

慢性疼痛試験で稀に用いられている優越性を証明するアプローチの 1 つは、さまざまな用量の実薬を比較して高用量での効果が有意に大きいことを証明するものである。このような「用量反応デザイン」は、低用量がプラセボよりも悪いと考える理由がない限り優越性および高用量と低用量の有効性の比較を証明するためにプラセボ群を必要としない[8,119]。さらに、このようなデザインでは全被験者が実薬を投与されることから薬剤の副作用により被験者および試験責任医師が非盲検化される可能性は低い。用量反応デザインのこのような側面は、プラセボを投与される可能性がないことから患者にとってより魅力的となり、組入れも促進する可能性がある。しかし、用量反応デザインでは有効に実施するために有効用量範囲の知見を必要とする。さらに、有効な治療薬の高用量と低用量を比較して優越性を証明することは、高用量の有効な治療薬とプラセボを比較して優越性を証明する場合よりも困難となる可能性が高い。

#### 6.5. 「上乗せ」試験

弱いレシキュー鎮痛薬(例 アセトアミノフェン)しか提供されない場合、慢性疼痛患者をプラセボ群に無作為割り付けすると問題であると考えられる場合がある。例えば、重度の発作性疼痛を伴った三叉神経痛患者をこのような試験に組入れることはおそらく倫理的とはみなされない。代替策としてはおそらく、有益ではあるが完全な鎮痛は得られていなかった標準治療(例 カルバマゼピン)を継続したまま被験薬またはプラセボに無作為割り付けすると思われる[154]。これはてんかんに関する RCT の標準デザインであるが、比較的少数の慢性疼痛に関する確認試験で同じデザインが利用されている。併用薬およびレシキュー薬に関するセクションで述べたように、このような「上乗せ」アプローチが利用される場合に併用鎮痛薬の使用を認めることが可能なくつかの異なるアプローチがあるが、慢性疼痛に関する確認試験におけるこれらの至適利用を促進するには追加的な評価を受けなければならない。

このような臨床試験デザインの考慮のための追加的な推進力となるのは、慢性疼痛に利用可能となりつつある薬剤数を増やすことである。このような薬剤に異なる作用機序がある場合、これらを併用することによって相加的または相乗的な作用がもたらされ、反応例の割合もしくは反応例におけるベネフィットの程度のいずれかを高める可能性がある[例 46,47]。

#### 6.6. その他の臨床試験手法

臨床試験のデザインおよび手法にはさまざまな進歩があり、ますます注目を集めている。これらには、Bayesian 法および適応的手法[13,45]、選好性研究[62,80]および無益の場合の早期中止がある。しかし、現時点でこれらのアプローチを用いた慢性疼痛に関する RCT は少なく、慢性疼痛に関する確認試験に対するそれらの長所および短所の評価の際の基礎となる情報は限られている。

#### 7. 結論および推奨事項

表 1~3 は慢性疼痛に関する確認臨床試験への推奨事項の要約を示したものである。繰り返しになるが、これらの推奨事項の多くは系統的研究に基づいておらず、むしろ慢性疼痛およびその他の疾患の治療に関する臨床試験の方法および結果からの外挿である点が重要である。これらの限界があることから、本稿の推奨事項は出版や助成金応募または規制当局による製品承認に関する要件と見なされてはならない。しかしながら、慢性疼痛に関する確認臨床試験の方法論的側面に対する注目および研究が増えればこれらの分析感度は高まり、最終的には慢性疼痛治療に関するより有意義な評価が得られるようになるだろう。

慢性疼痛に関する確認臨床試験に関して、以下の研究アウトカムに対する影響を含め、本稿の範囲を超えた研究デザイン上の検討事項は多数ある：(1) 研究実施施設数[33,69]；(2) 採用率[61]；(3) 試験責任医師、その他の研究担当者および被験者の訓練[22,33,79,158]；(4) アウトカム指標に紙による記録を用いているか電子的記録を用いているか[84]；ならびに(5) データ収集の頻度[64]。急性疼痛および慢性疼痛に関する臨床試験の実施に伴うこれらをはじめとする実用的側面は"Good Research Practice"と呼ぶことができ、近く発表される IMMPACT の推奨事項において検討されるだろう。

否定的な結果が出たり失敗した慢性疼痛試験についての最近の複数の報告では著者らのレビューおよび推奨事項に対する推進力が得られたが、これらの不成功に終わったアウトカムに対する説明では、これらの試験デザインにおける欠点のみが関係しているわけではない可能性を認識することが重要である。むしろ、慢性疼痛の原因に関する知識の欠如ならびに病態生理学的機序および治療反応における患者の不均一性への理解不足が試験の不成功の主な説明となっている可能性がある。慢性疼痛の病態生理への関心は非常に高まっており、現在進行中の研究の焦点は、さまざまな症状の機序および個々の患者においてこれらの機序を確実に特定するデバイス手法の特定、ならびにこれらの機序を標的とした治療の開発に合わせられている。これらの取り組みの最終的な目標は、治療介入が患者の疼痛の特定の機序を標的とする、機序に基づいた治療アプローチの基盤とすることである[9,65,92,126,151,152]。慢性疼痛に関してこのような機序に基づいた理解が深まれば、臨床試験における適格基準には違った定義がなされ、プラセボとの比較における治療効果の証明に対する期待は変わると思われる。

また、臨床試験には、少なくとも現行の形態においては、慢性疼痛症状に対する有効な治療介入の特定に関して大きな限界があるという可能性を考慮することも重要である[62]。この観点から言うと、否定的な結果が出たり失敗した慢性疼痛試験は、慢性疼痛の複雑性にもっと上手く対処できる可能

性がある試験デザインで革新への推進力を提供している。

Rowbotham[117]は、臨床試験結果を公表しなかったことによって生じる疼痛治療理解への悪影響について最近記述している。肯定的か否定的かを問わず、すべての試験結果の公表を求めた Rowbotham の言葉をそのまま繰り返すことによって本稿の結びとしたい。慢性疼痛に関する確認試験のデザインにおける改善は、これらの臨床試験の方法論的特徴およびアウトカムを記述した一連の包括的な情報が利用可能な場合のみ実現できる。

#### 謝辞

本稿に示した見解は著者らのものであり、著者の中に本稿で述べた事項に関連して金銭的な利益相反を有するものはない。米国退役軍人局、米国食品医薬品局、米国国立衛生研究所または使途に制限のない助成金を Rochester 大学継続専門教育事務局に提供した製薬企業からの公式な承認を意味するものではない。

著者らは、IMMPACT 会議の組織に貴重なご支援をいただいた Paul J. Lambiase および Mary Gleichauf に感謝申し上げます。また、"Good Research Practice"の用語をご提案いただいた NPK にも感謝申し上げます。本稿の元となったコンセンサス会議を含む IMMPACT の活動の支援を目的として、Allergan、Alpharma、AstraZeneca、Celgene、Cephalon、Eli Lilly、Endo、GlaxoSmithKline、Johnson & Johnson、Merck、NeurogesX、Pfizer および Schwarz Biosciences より使途に制限のない助成金が Rochester 大学継続専門教育事務局に提供されました。

図 1. 慢性疼痛に関する確認試験のための並行群間研究デザイン

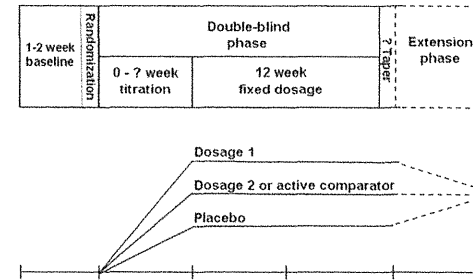


Fig. 1. Parallel groups research design for confirmatory chronic pain clinical trials.



表 1. 慢性疼痛に関する確認試験のための研究デザインの検討および推奨事項:

研究被験者

1. 人口統計学的特徴および関連するその他の特徴

- ・治療の有効性、安全性および忍容性が患者背景に応じて変化するか否かの研究が可能となるような重要な人口統計学的情報(例 年齢、性別、人種/民族)を捕捉する。
- ・基礎疾患、疼痛状況および治療介入に関連した重要情報を捕捉する。
- ・研究結果の一般化可能性の評価に必要とされる追加情報があれば捕捉する。

2. 診断

- ・診断の基準および手順を標準化する。

3. 罹病期間および疼痛持続期間

- ・疼痛の自然消散の可能性を限定する最小罹病期間を明確にする。
- ・最小疼痛持続期間は慢性疼痛の場合 3 ヶ月とすべきであるが、6 ヶ月も検討可能である。
- ・罹病期間が長すぎる患者については、治療への反応の可能性が低いと考えられる理由がある場合には除外を検討する。

4. 疼痛強度

- ・ベースライン時の患者の自己申告による最小疼痛強度は、ほとんどの場合において 0~10 の数値化スケールにおいて 4 または 5 に相当するものであること。
- ・ベースライン時疼痛強度が極めて強い患者およびベースライン時の疼痛の変動が激しい患者については除外を検討する。

5. 内科的または精神科的な併存疾患

- ・標的疼痛と同一部位により強い疼痛を有する患者または標的疼痛と明確に識別できない疼痛を有する患者は除外する。
- ・標的疼痛を増強する可能性があるか、さもなければ標的疼痛の評価に影響を及ぼす可能性がある内科的および精神科的な併存疾患を有する患者は除外する。
- ・治療が重篤な有害事象を引き起こすと考えられる可能性がある内科的または精神科的な併存疾患を有する患者は除外する。

6. 鎮痛に関する併用薬およびレスキュー薬ならびにその他の治療

- ・試験期間中はいかなる薬物療法または非薬物療法(例 理学療法、鍼)も開始または中止すべきではない。
- ・鎮痛に対する併用薬が禁止される場合: (1)十分な休薬期間を設けなければならない、また、(2)プラセボ群を置いた試験では、レスキュー鎮痛薬を提供すべきであるとする倫理的要件がある\*。
- ・併用鎮痛薬の使用が認められる場合、通常は試験期間を通して安定用量を維持すべきであり、プロトコールには被験者が以下の継続を認められるか否かを事前に設定しておくべきである: (1)あらゆる鎮痛薬; (2)研究対象の慢性疼痛に対し国または地域当局から承認された 1 種類以上の薬剤(「上乗せ」デザイン); または(3)それぞれについて最高用量が定められた、限定された数の特定の薬剤。

7. 前治療に対する反応

・有効性の欠如または忍容性の不良により複数の前治療が不奏効であった患者については、新規作用機序を有する薬剤がそのような患者に有効となる可能性があることを認識することは重要であるが除外を検討する。

注: 本表の推奨事項は特定の目的および慢性疼痛に関する確認臨床試験に利用可能な研究デザインのすべてに必ずしも当てはまるものではなく、出版や助成金応募または規制当局による製品承認に関する要件と見なされてはならない。\*例外としては、主要アウトカムが耐え難い疼痛による中止までの時間である場合が考えられる。

表 2. 慢性疼痛に関する確認試験のための研究デザインの検討および推奨事項:

#### 臨床試験の相

##### 1. ベースライン期間

- ・研究手順へのアドヒアランスが不良の被験者(例 最低日数分の疼痛日誌の記入を完了しない被験者)を特定し検討する。
- ・ベースライン期間の最短期間は通常 1 週間とすべきで、特定の条件下では、2 週間のベースライン期間を設定すると費用がかさみその他の不利益が生じる可能性があったとしても潜在的ベネフィットはそれらを上回るものとなる可能性がある。
- ・ベースライン期間中のレスキュー鎮痛薬の使用許可を検討する。

##### 2. プラセボ導入期間

- ・目的が分析感度の強化である場合、大うつ病に関する臨床試験に基づけば単盲検によるプラセボ導入期間の利用がその目的を達成することを示唆する証拠はほとんどないことから、単盲検によるプラセボ導入期間は一般に推奨されない。その他に考えられる単盲検によるプラセボ導入期間の目的については検討の価値がある。
- ・二重盲検によるさまざまな期間を用いたプラセボ導入期間などの代替案を検討すべきである。

##### 3. 無作為割り付け前の強化

- ・それぞれの試験に特有の状況に応じて、強化戦略は実薬治療に対する反応または忍容性を高め、最終的に試験から脱落する被験者数を抑える可能性があるものとして考慮することができる。
- ・強化は、明確な疼痛機序もしくは薬物動態または薬力学における個人差の結果として潜在的奏効例からなる患者サブグループが存在する場合に特に有用となるかもしれない。
- ・強化が試験結果を目的とする集団への一般化可能性に及ぼす影響を考慮すべきである。
- ・無作為化前に実薬治療の副作用について被験者に気付かせることによって盲検化に支障をきたす可能性がある点に考慮すべきである。
- ・被験薬の奏効例を選択することによる強化は、先行曝露による薬力学的効果が長時間作用性である場合は適切でない可能性がある。

##### 4. 漸増相、維持相、漸減相

- ・副作用の発現率および重症度を低減し、特定の有害事象の発現を抑える必要がある場合には漸増相が推奨される。その実施期間は目的達成に必要なだけの短期間とすべきである。
- ・12 週間の維持相は、特定の状況から短期間または長期間が必要とされる場合を除き、慢性疼痛治療に関する確認試験では一般に推奨される。
- ・漸減相は、突然の治療中止が退薬症状や中断症候群を引き起こす可能性がある場合は一般に推奨されるが、その期間はそれらを予防または最小限に抑えるのに十分な長さとするべきである。特定の条件下では、突然の薬剤中止が退薬症状の調査に適している場合があるかもしれない。

注: 本表の推奨事項は特定の目的および慢性疼痛に関する確認臨床試験に利用可能な研究デザインのすべてに必ずしも当てはまるものではなく、出版や助成金応募または規制当局による製品承認に関する要件と見なされてはならない。

表 3. 慢性疼痛に関する確認試験のための研究デザインの検討および推奨事項:

#### 治療群

##### 1. 治療群の数

- ・治療群の数は臨床試験の目的を鑑みて可能な限り少数にとどめるべきで、一般に 3 群を上回らないようにすべきである。例えば、3 群では、被験薬、プラセボおよび第 2 の用量の被験薬または実薬対照を投与する。

##### 2. プラセボ群

- ・有効性が確立されている既存薬はプラセボに対して常に優越性を示すわけではないため、慢性疼痛についての研究ではプラセボ群を置くことが一般に推奨される。
- ・薬剤の副作用によって盲検解除が起こる可能性があることから、試験終了時に被験者と試験責任医師が被験者が割り付けられていた治療群についてどのように推測しているかおよびその推測に至った主な理由について評価するよう検討すべきである。
- ・アクティブプラセボは副作用による盲検解除を低減できるが、有害事象および中止例の発現率が上昇することやベネフィットはもたらさないが潜在的悪影響があると考えられる薬物に被験者を曝露させることに関する倫理的懸念があることから、アクティブプラセボの使用はほとんどの場合で推奨されない。

##### 3. 固定用量戦略対可変用量戦略

- ・有効用量を確立するには固定用量群が一般に推奨される。
- ・可変用量群は固定用量群に比べて分析感度が高い可能性があるだけでなく、臨床診療への一般化可能性が高くなる可能性がある結果が得られることが考えられ、多くの場合において、固定用量に基づく有効性および安全性のデータを補完する有用な情報を提供するものとして可変用量の評価は推奨できる。

##### 4. 実薬対照

- ・リソース面から追加的な治療群を置くことが許され、そのことが研究の実行可能性または分析感度に影響を及ぼす可能性があるレベルにまで試験群の数を増やすようなことにならない場合には、有効性が十分に確立された実薬対照を含めることが推奨される。
- ・プラセボ群が置かれていない非劣性試験または同等性試験は有効性を十分に証明できず、プラセボを置かない実薬対照優越性試験が必要である。

注: 本表の推奨事項は特定の目的および慢性疼痛に関する確認臨床試験に利用可能な研究デザインのすべてに必ずしも当てはまるものではなく、出版や助成金応募または規制当局による製品承認に関する要件と見なされてはならない。

