

背景

ほとんどとは言えないにしろ、拘禁されている犯罪者の多くは、物質乱用の問題を抱えている。効果的な治療を行わなければ、これらの物質乱用犯罪者は、薬物以外の犯罪をも何度も繰り返すようになりやすい。拘禁期間は、薬物乱用と犯罪の悪循環を断ち切るよい機会となる。拘禁下では多くの種類の薬物乱用治療が行われているが（例：治療共同体、集団カウンセリング）、これらのプログラムの効果は不明である。

目的

本研究の目的は、拘禁下で行われた薬物治療プログラムが、釈放後の再犯と薬物再使用を抑制する効果についての準実験および実験（RCT）による評価研究を系統的にレビューすることである。第二の目的は、プログラム、サンプル、研究法によって効果の違いがどれほどあるかを調べることである。新たな研究を用いてレビューを最新のものにするために、2006年のオリジナルのレビュー（Mitchell, Wilson, and MacKenzie, 2006を参照）のアップデート版であるこのレビューには、オリジナルのレビュー以後に発表された研究を新たに含めている。

検索方法

レビューに含める候補となる研究を見つけるために、文献データベースの検索、選択した学術誌のハンドサーチ、薬物治療研究を行っているいくつかの研究機関のウェブサイトの閲覧を行った。

検索基準

選択候補の研究として必要な基準は、拘禁下（例：ジェイル、刑務所）で行われる薬物治療プログラムの効果を評定していること、実験的または準実験的な対照群を有した研究デザインであること、釈放後の再犯または薬物使用アウトカムを測定していること、1980年から2011年までに実施されたものであること、である。

データ収集と解析

それぞれの評価研究から、再犯や薬物再使用のさまざまな測度に関する各プログラムの効果を定量化するために効果量をコード化した。さらに、プログラムの特性、研究方法、サンプルもコード化した。効果量は、分散の逆数で重みづけされたランダム効果モデルのメタ・アナリシスで解析した。

主たる結果

74の研究がわれわれの選択基準に合致した。これらのプログラムの全般的な平均効果量は、再犯と薬物再使用をおよそ15から17%減少させるものであった。しかし、これらのプログラムの効果は、プログラムのタイプによってさまざまであった。治療共同体が再犯および薬物再使用を抑制する効果は、比較的一貫していたが中程度の効果であった。カウンセリングと薬物維持プログラムの効果はまちまちであった。具体的には、カウンセリング・プログラムは、再犯を抑制したが、薬物再使用には効果がなかった。薬物維持プログラムは、薬物再使用の抑制にはかなり大きな効果があったが、再犯には効果がなかった。一方、ブートキャンプの再犯と薬物再使用抑制効果は、無視できるほどのものしかなかった。

結論

拘禁下における薬物治療プログラムの効果を評価したこのレビューでは、これらのプログラムには再犯を抑制する中程度の効果があることを見出した。その知見は、治療共同体の効果を強く支持しており、それはこの種のプログラムが、比較的一貫して再犯と薬物再使用を抑制していることによる。カウンセリングと拘禁下の薬物維持プログラムの効果は、まちまちであった。カウンセリング・プログラムは、再犯の低下とは関連していたが、薬物再使用とは関連がなかった。一方、拘禁下における薬物維持プログラムは、薬物再使用の低下とは関連があったが、再犯とは関連がなかった。ここで注意すべき点は、拘禁下における薬物維持プログラムの効果に関するわれわれの知見は、コミュニティにおける薬物維持プログラムに関する大規模なレビューの結果とは異なっているということである (Egli, Pina, Christensen, Aebi, and Killias, 2009w を参照)。最後に、薬物犯罪者に対するブートキャンプ・プログラムは、再犯と薬物再使用に対して、無視できるほどの効果しかなかった。

一般用語での要約

この研究は、メタ・アナリシスの手法を用いて、拘禁下で行われた薬物治療プログラムを評価した 74 研究の結果を統合したものである。拘禁下で行われる薬物治療プログラムは、4 つのそれぞれ異なったタイプに分類される。それは、治療共同体 (TC) 、 集団カウンセリング、薬物犯罪者に特化したブートキャンプ、そして薬物維持プログラムである。われわれは、これらの各タイプのプログラムが釈放後の再犯や薬物使用を抑制する効果を調べた。さらに、研究結果の違いが、研究方法、サンプル、プログラムの特質の違いによって説明でされ得るのかどうかについても検討した。われわれの研究の結果は、TC プログラムが両方のアウトカム指標において効果があることを一貫して支持していた。そして、この知見は、研究方法、サンプル、プログラムの特質の違いにかかわらず確固たるものであった。また、集団カウンセリング・プログラムが、再犯を抑制する効果も支持された。しかし、これらのプログラムが薬物使用に対する効果は無視できるほどのものでしかなかった。薬物維持プログラムの効果もまた、薬物使用の抑制には効果があった一方で、再犯には効果がなかった。ブートキャンプには、どちらのアウトカム指標にも大きな効果はなかった。

1. レビューの背景

研究によれば、拘禁されている犯罪者のうち、かなりの部分が薬物依存者であることがわかっている。たとえば、Peters ら (Peters, Greenbaum, Edens, Carter and Ortiz, 1998) は、テキサスの受刑者サンプルのうち、56%の者が収容される前30日間に物質乱用か物質依存障害の診断を受けていたと報告している。同様に、オハイオの ジェイル受刑者のうち、51%が現に薬物依存であった (Lo and Stephens, 2000)。国全体を代表するサンプルに対するより最近の調査でも、これらの知見が支持されている。2002年に実施されたジェイル入所者の国全体を代表するサンプルに対する調査では、36%が薬物依存であり、18%が薬物を乱用していることがわかった (Karberg and James, 2005)。これらの推計にアルコールを含めると、ジェイル入所者の68%が依存 (45%) または乱用 (23%) ということがわかった。同様に、州立および連邦刑務所の受刑者に対する2004年

の調査では、州立刑務所受刑者の53%、連邦刑務所受刑者の45%が、DSM-IVの薬物依存または乱用の基準に合致することが見出された(Mumola and Karberg, 2006)。実際、明白に薬物乱用治療が必要なアメリカ人全体のおよそ40%が刑事司法システムの指導下にあると推定されている(Gerstein and Harwood, 1990:7)。

拘禁下にある犯罪者の薬物依存を調査した国外の研究はずっと少ないが、合衆国以外の他の多くの国においても、薬物依存は拘禁下にある犯罪者の間によくみられる問題であるようだ。たとえば、Bennett (1998)は、イギリスの5つの都市で逮捕された者のサンプルのうち、45%がこれまでの生涯のある時点で薬物依存であり、33%が現に薬物依存であったと報告している。同様に、カナダの連邦刑務所に収容されている受刑者のうちの31%、州刑務所に収容されている受刑者のうちの43%が薬物依存であった(Pernanen, Cousineau, Brochu, and Sun, 2002)。

効果的な薬物乱用治療がなされなければ、これらの薬物依存受刑者のうちのきわめて多くの者が犯罪を続けることになるであろう。実際、司法省統計によれば、保護観察を受けた者のうち、深刻な薬物依存者は、薬物依存でない者よりも再逮捕される率が53%高い(Bureau of Justice Statistics 1995: 26)。このような意味で、犯罪者が刑務所に収容されている時期は、薬物依存と犯罪の悪循環に介入して、犯罪を防止する重要な機会である。

矯正施設(刑務所やジェイルなど)の有するいくつかの性格ゆえ、拘禁下における薬物乱用治療は望ましいものとなっている。第一に、社会内におけるよりも、薬物を入手できる機会がずっと限定されており、それによって解毒や治療中の断薬が促進できること。第二に、治療や内省に集中できる時間が豊富にあること。そして、おそらく最も重要な点は、矯正施設は、そうでもしなければ治療を受けようとしめないような薬物乱用受刑者に対し、相当な強制力を持って治療を受けるように働きかけることができるという点である。

拘禁下の薬物乱用治療は実に多様であり、集団・個人精神療法、12ステップ・グループ、メサドン維持療法、さらには薬物乱用者向けブート・キャンプなどの広範な種類のもが含まれている。したがってわれわれはこのレビューのため、これらのプログラムを、薬物乱用者(すなわち、違法薬物、および/またはア

ルコール)を対象とし、薬物乱用と他の犯罪行動を減少させることを目的とし、それらの介入は矯正施設内において実施されるものである、定義とした。現在実施されている拘禁下における薬物乱用治療プログラムの評価は、治療共同体(TC)、グループ・カウンセリング・プログラム(薬害教育、AA/NAなどの12ステップ・グループ)の効果の測定に焦点を当てたものが大部分である。薬物乱用者を特に対象としたブート・キャンプや、薬物維持療法プログラムの効果を査定する評価研究はきわめて少ない。

TCの個々のプログラム要素はまた、プログラムによって大きく異なっている。しかし、いくつかの要素は共通である。まず、社会復帰に役立つような雰囲気醸成するために、治療共同体のメンバー(residents)は、通常はそれに参加していない受刑者とは離れた独立の治療区画に分離して収容される。第二に、治療スタッフの指導と監督下にあるメンバーは、治療セッションの主宰、他のメンバーがルールを遵守しているかどうかの監視、治療区画の清掃、仲間内の問題解決などに補助的にかかわる。第三に、TCのスタッフとメンバーは、ルール違反があったときは、対立することもあるが、同時にメンバー同士はまた、望ましい変化を成し遂げるための互いの努力をサポートする。第四に、TCの運営方針は、薬物乱用はより広汎な個人的障害の1つの症状であり、治療の焦点は薬物乱用そのものではなく、その背後にある障害や心理的問題に置かれるべきだという考えに基づいている。

カウンセリング・プログラムは、定義づけすることがいくぶん困難である。一般的に、これらのプログラムには、グループ・カウンセリング・プログラム(例:AA, NAなどの12ステップ・グループ)、生活技能訓練、認知的スキル訓練、薬害教育、成人のための基礎学力の教育などが含まれる。これらのプログラムに共通する主要な内容は、グループに依拠した治療への信頼であり、そこでは物質乱用と他の一般的な問題をお互いの問題として解決するために、仲間内で話し合っていこうとする。しかし、すべてのカウンセリング・プログラムが仲間同士の治療に頼っているわけではなく、クライアントと治療者が薬物問題の治療を目指して共に取り組むような個人ベースのものもある。さらに、グループ・カウンセリングと個人カウンセリングの双方を含むようなカウンセリング・プログラムもある。

ブートキャンプは、軍隊の基礎訓練をモデルとしたものである。受刑者は、厳しい運動訓練に参加し、軍隊式訓練や儀礼の学習、制服の着用、障害物コースへの挑戦(時間内に障害物コースを走破)などを行う。ブート・キャンプは高度に組織化されている。参加者は、起床してから消灯するまで、常にスケジュールに沿った行動に従事させられる。ブート・キャンプにはまた数多くの対立場面があるが、ほとんどのTCとは違って、対立場面の多くが矯正スタッフと受刑者の間に生じるもので、訓練教官が定められた行動様式からの逸脱に対して統制を行うものである。理論上では、苛酷で厳格なブート・キャンプの性格は、将来の犯罪行為を抑止するものであり、これらのプログラム内容はプログラム参加者の中に自己統制を植えつけ、再犯の減少を導くものである (Wilson and MacKenzie, 2006)。

薬物維持療法(Narcotic maintenance program)(メサドンやレボ・アルファ・アセチル・メタドール[LAAM]などによる)は、他の拘禁下における薬物治療プログラムとは非常に異なっている¹。これらのプログラムは、合成オピオイドを処方することによって、ヘロイン依存に伴う害(たとえば、感染症、犯罪行為)を減少させようとするものである。ヘロインとは異なり、これらの薬物は高揚感を生じさせることはなく、その代わりに、アヘン系薬物による高揚感を遮断し、離脱症状を抑制する。これら薬物維持療法の中には、アヘン系薬物への依存が改善されるまで患者に対して処方する薬物の量を徐々に減少していく場合もあれば、一生薬物を処方し続ける場合もある。

これら上述したどのタイプの薬物乱用治療プログラムも、薬物乱用と他の犯罪行為を減少させる明確な可能性を有している。しかしながら、一連の文献の現存するレビューによれば、TCプログラムの効果を支持する強いエビデンスがあったのみであった (Pearson and Lipton, 1999)。特に、Pearson and Lipton (1999) は、拘禁下における薬物乱用治療プログラムの再犯減少における効果を評価する研究を系統的にレビューした。彼らの系統的なレビューは、拘禁下(「刑務所、拘置所、ジェイル、または身柄を収容する同様の矯正施設」(p390))でさまざまな国において、1968年から1996年までの間に終結した介入を、準実験的及び実験的に評価したものを包括的に検索したものである。その検索によ

¹ LAAM は、米国ではもう使用されていないが、Kinlock, Battjes, Schwartz, and the MTC Project Team (2005)において使用された薬物である。

れば、30の研究が彼らの選択基準に合致した。Pearson and Liptonがこれら30の研究を統合した結果、TCは再犯を減少させる効果があったということである。特に、彼らが分析したところでは、レビューされた7つのTC研究のうち、6つが、再犯防止に著明な効果があり、重みづけられたr効果量の平均は、0.133 ($p = 0.025$) で、正の効果量のレンジは.13から.28 までであった(効果量が負(-0.16)であるものが1つあった)。これに対して、ブート・キャンプやグループ・カウンセリング・プログラムでは、平均効果量は、統計的に有意なものはなく、これらのプログラムは、プログラムを何も実施しなかったときに比べて効果があるとは言えなかった。さらに、Pearson and Liptonによれば、他のタイプの介入については、評価研究が少なすぎて、その効果についてははっきりとした結論を導き出すことができなかった。しかしながら一般的に見て、彼らはメサドン維持療法、薬害教育、認知行動的プログラム、12ステップ・グループの効果を評価するエビデンスについて、効果が期待できるものであったと述べている。

多くの点において、われわれのこの系統的レビューは、Pearson and Liptonのレビューを拡大したものである。Pearson and Liptonの研究のように、このレビューは、拘禁下における薬物乱用治療の治療後の薬物使用と他の犯罪行動についての効果を、メタ・アナリシスの手法で組織的・包括的にレビューしたものの総括である。彼らの研究とこの系統的レビューの主要な違いは、本研究がもっと最近の研究(1980年から2004年)を取り扱っているという点である。この違いは、次の2点で重要なものとなる。(1)薬物乱用治療の最近の評価研究の方が、現行の矯正実務についてより一般化でき得る。(2) 拘禁下の薬物乱用治療プログラムについての多くの評価研究は、1996年以降になされた。これらの時間枠についての相違を考慮すれば、われわれの結果は Pearson and Liptonの研究よりいくぶん異なったものとなるであろう。

2. レビューの目的

このレビューの目的は、拘禁下における薬物乱用治療への介入が薬物のリラプスと再犯を減少させることに対する効果についての利用可能なエビデンスを系統的に総括することである。より具体的には、この系統的レビューは以下のようなリサーチ・クエスチョンに答えようとするものである。拘禁下における

薬物乱用治療プログラムは、再犯と薬物使用を減少させる効果があるのか。これらのプログラムにはおおよそどの程度の効果があるのか（たとえば、効果の大きさは？）。特に効果的である、または効果のないプログラムとはどのようなタイプのものか。効果的なプログラムと効果のないプログラムの違いはどのようなようなものか。これらの疑問は、数量的なメタ・アナリシスの手法によって回答できるものである。

3. 方法

3.1 レビューに含むべき研究と除外すべき研究の基準

このレビューが扱う範囲は、少年と成人に対する拘禁下における薬物乱用治療プログラムに対する対照群を用いた実験的・準実験的評価研究である。本レビューの適格基準は次のようなものである。(1) 矯正施設（刑務所、ジェイルなど）において実施された介入を評価する研究、(2) 物質乱用者に特に対象を絞った介入、(3) 治療を行わないか、最低限の治療しか行わない対照群を有する実験的、または2群の準実験的研究デザインを利用した評価研究、(4) 釈放後の犯罪行為（薬物使用を含む）を含む効果の測定結果を報告している、(5) 介入は、1980年から2011年11月の間に実施されている、(6) 研究は、効果量を計算するために十分な情報を報告していること、また、対象となる研究は、公表されたものでもされていないものでも差支えない。

一番目の基準について、われわれの「矯正施設」についての操作的定義は、ジェイルと刑務所、そして少年用の同様の施設のみに限定している。ーフウェイ・ハウスや社会内の収容施設で実施される介入は含んでいない。さらに、この基準では注目すべきいくつかの研究が除外されていることを述べておく。特に、Dynia and Sung (2000) や Knight and Hiller (1997) が報告したような拘禁に代わる代替法として考案されたプログラムは、この基準によって除外された。

第二の基準は、このレビューの焦点を、特に薬物乱用者を対象とした研究に限定するものである。矯正施設において実施される治療的介入のうち、薬物使用歴に関係なく、一般的に受刑者に対して実施されるようなものは含まれてい

ない。たとえば、Shaw and MacKenzie (1990) はブート・キャンプ・プログラムの薬物乱用受刑者サブ・グループへに対する効果を評価しているが、この評価研究は対象から除外された。というのは、このブート・キャンプ・プログラムは、薬物乱用者のみを特定して対象としたものではないからである。同様に、Jones Olson Karr, and Urbas (2003 ほか) は、イリノイ州における矯正ブートキャンプの効果を毎年報告しているが、これもまた薬物乱用者を特定して対象としたものではなく、本研究からは除外された。これとは対照的に、Zhang (2000) は、薬物乱用者に限定したブートキャンプ・プログラムの評価をしており、この評価研究はこのレビューの対象となっている。この基準が必要であるのは、このレビューは、拘禁下における薬物乱用治療を対象としているからである。この基準がなければ、このレビューは、薬物乱用者を含む拘禁下における介入のレビューとなってしまふ（さらに、受刑者の大多数が薬物乱用者であることを考えると、そのようなレビューはほとんどすべての拘禁下における介入のレビューとなる危険を冒すこととなってしまふ）。

第三の基準は、対象となる評価研究はすべて、治療を受けていないか、あるいは最低限の治療しか受けていない比較・対照群を有するものでなければならないということを述べたものである。したがって、大まかな類似性のある介入や、再犯を減少させる効果の比較が論議を呼んでいるような2つ以上の介入法を比較している準実験的な研究（たとえば、ある治療法と別の治療法の比較、治療回数の違いによる変化の比較など）は除外した。さらに、対照群を欠いた準実験デザインの研究も除外した。この基準に従って、Sacks, Sacks, McKendrick, Banks, and Stommel (2004) やSullivan, MaKenrick, Sacks, and Sacksの研究は、比較群が十分な治療サービスを受けていたため除外した。さらに、対照群が主として、またはもっぱらプログラム脱落者から構成されているような評価研究も対象とはしていない。たとえば、Field (1985, 1989) や Berggren and Svard (1990) の研究では、プログラムから脱落した者を対照群として利用しており、本系統的レビューからは除外した。というのも、これまでの研究が明白に示しているように、薬物乱用治療プログラムからの脱落者と、プログラム修了者は、介入の前から存する重要な客観的変数について（さらには、目に見えない重要な他の変数についても）異質であることが多く（たとえば、Hiller, Knight, Simpson, 1999参照）、それゆえに、このような研究デザインでは、選択バイアスがとりわけ大きな問題となるからである。

第四と第五の基準は、ほとんど説明を要さないであろう。しかし、強調しておきたいのは、すべての研究は、釈放後の再犯についての情報が必要だということである。この基準によって、Shewan, Macpherson, Reid, Davies (1996) や Dolan, Shearer, MacDonald, Mattick, Hall, Wodak (2003) の研究のような刑務所内での治療結果を示したのみの少数の研究は、優れた研究であるが除外した。1980年より前に実施された研究も、現在の矯正に対する一般化可能性を高めるために除外した。

最後の条件は、効果量を計算するのに必要な情報が報告されていない研究を除外するものである。この基準は、実務的な目的から必要なものである。残念なことに、これ以外の基準は満たしていただきたいいくつかの研究が（たとえば、Schippers, Van Den Hurk, Breteler, Meerkerk, 1998; Guerin, 2002）、この最後の基準によって除外された。

3.2 適格研究の検索方法

検索の目的は、上記の選択基準に合致するすべての研究を、公表されたものであろうとそうでなかろうと、全部見つけ出すことである。この目的を達するために、まず文献データベースのコンピュータによるキーワード検索を行った。われわれの最初のレビューでは、次のようなデータベースでの検索を行った。すなわち、PsychLit, MedLine, NCJRS, Criminal Justice Abstracts, Dissertation Abstracts, Sociological Abstracts, Social Science Citation Index, SocioFile, Conference Papers Index, UnCover, C2 SPECTR, CINAHL, そしてグーグルによるインターネット検索である。このアップデート版のレビューでは、基本的に同じデータベースを検索したが、例外としてわれわれが使用できなくなった少数のデータベースを除外し。具体的には、アップデート版で検索したデータベースは、PsychInfo, MedLine, NCJRS, Criminal Justice Abstracts, Social Science Full Text, ProQuest Dissertations and Theses, Sociological Abstracts, Conference Papers Index, IngentaConnect, CINAHL, グーグルである。いずれの検索でも以下のキーワードを用いた。薬物乱用治療, 物質乱用治療, 薬物カウンセリング, 治療共同体, メサドン維持療法, ブート・キャンプ, 犯罪者, 入所型物質乱用治療 (RSAT), RSAT, 飲酒運転, DUI,

DWI, 受刑者, 拘禁, 拘禁された, 刑務所, 評価, 効果の評価, 再犯。これらのキーワードが, さまざまな組み合わせで用いられた。どのような組み合わせを用いたかの詳細については, 付録-1を参照されたい。

さらに, 未知の研究を探すために, 既存の研究論文や文献レビューを注意深く読んで適格研究を検索した。特に, 適格研究を見つけるために, 現存する論文の参考文献リストを調べた。同様に, 適格となった研究の多くは, 同様の研究を参照していたので, これらの研究の適格性もまた調べた。加えて, 「刑務所における薬物使用とHIV/AIDSについての研究ダイジェスト」(Flanagan, Arsovska, Giaime, Goril, Kahl, Król, Moore, 2004) を調べた。これは, 「灰色」研究, 特にヨーロッパにおける「灰色」の研究の要旨の多くが掲載されている。特に, これには複数言語で出版された研究が含まれている。

さらに, いくつかの主要な研究機関のウェブサイトも検索した。特に, 以下のようなサイトの該当する研究レポートを検索した。カナダ矯正局研究出版のページ (http://www.csc-scc.gc.ca/text/research_e.shtml), 英国内務省 (<http://homeoffice.gov.uk/>), RAND薬物政策研究センター (<http://www.rand.org/mult/dprc/>), 都市犯罪・司法研究センター (<http://www.urbenorg/justice/indez.cfm>), Vera司法出版所 (<http://vera.org/publications/publications.asp>)。

われわれはまた, 1999年から2011年11月の間に, 以下の学術誌に掲載された論文のタイトルと要旨を手作業で検索した。すなわち, Journal of Substance Abuse Treatment, International Journal of Offender Therapy and Comparative Criminology, Journal of Drug Issues, The Prison Journal, Crime & Delinquency, Journal of Offender Rehabilitation である。われわれがこれらの学術誌を手作業で調べることにしたのは, 該当する研究の厳密な発刊記録があることや, これらの学術誌の多くは, われわれが利用したコンピュータ・データベースにすべて掲載されているとは限らないからである。

最後に, われわれの検索方法と検索結果を情報専門家に見てもらった。情報専門家は, これらに加えて該当すると思われる研究の追加リストを提供してくれた。

タイトルとアブストラクトによる予備的審査によって該当可能性があると考えられたすべての研究について、最終的に該当するかどうかを決定するために綿密に検討した。特に、非該当であるとの明白な証拠がある各検索結果のタイトルとアブストラクトを調べた。すなわち、当該研究が1つ以上の選択基準(3.1 参照)に外れている明白な証拠を探して、各タイトルとアブストラクトを精査した。たとえば、評価されているすべての介入は矯正施設において実施されていることというのが第一の基準であるが、タイトルとアブストラクトから、矯正施設内で実施されているのではないプログラムが評価されているということが明白にわかる場合は、その研究は更なる検討からは除外された。タイトルとアブストラクトの調査では非該当であると除外できなかった研究を、さらに該当性の検討のために入手した。入手した研究は、注意深く読んだ上で、最終的に該当するかどうかを決定した。

3.3 適格研究で使用されている方法

適格研究の基本的研究デザインは、治療群と対照群を有し、釈放後の犯罪行為や薬物使用など、釈放後の治療効果についての情報を有するものである。対照群の構成法は異なっているが、共通するバリエーションとしては、時間経過による比較、近隣地区との比較、治療プログラムの受講資格があるが参加しないことを選択した受刑者、受講資格があるが薬物乱用治療プログラムの受講可能人員に限りがあるので受講できなかった受刑者、ランダム割付けなどが挙げられる。それぞれの研究はまた、選択バイアスの脅威を低下させるために適用した統計的方法の種類も異なっている(マッチング、共変量解析など)。対象となった研究はまた、再犯と認定する基準(たとえば、逮捕、起訴、再入)や、薬物を再使用したと認定する基準(たとえば、自己報告、尿検査)も異なっている。われわれは、これらの方法論的な差異を把握できるようにコード化した。

3.4 独立した研究か否かの判断基準

拘禁下における薬物乱用治療プログラムの評価においては、さまざまなタイプの統計的従属性が明らかとなった。よく見られた従属性は、犯罪行動を複数の方法で測定することや(たとえば、再逮捕、再起訴、薬物使用)、犯罪行動の同一の指標に対して異なるフォローアップ期間を設定することによって生じたもの

である。ほかの従属性は、同一の研究対象サンプルに対して複数の研究を実施し、その結果を報告したことによって生じたものである。

以下に詳細に述べる統計的方法には、研究結果の統計的独立性が要求される。われわれは研究結果の統計的独立性を担保するために、いくつかの方策を講じた。第一に、プログラムの効果を分析するに当たって、同一の評価の結果を報告している複数の研究を使って、別々の評価であるとするのがいかなる分析においても生じないことを確実にするため、すべての評価研究（治療群と対照群の比較など）を相互に1つ1つ精査した。第二に、犯罪行動の複数の測定値を報告している評価研究においては、これら複数の測定値を平均するのではなく、一連の選択基準を適用して、5つのデータセットを抽出し、その評価研究が各データセットに対してただ1つの効果量を提供するようにした。そして、次のような基準を満たす効果量を一番望ましいものとした。すなわち、(1) 一般的である（たとえば、特定の罪種でなく、全罪種をカバーするようなもの）、(2) 逮捕を基準としている、(3) 二値変数である、(4) サンプルのメンバーを12か月追跡調査している。これらの基準に合致する効果量を望ましいものとしたのは、逮捕は他のアウトカムよりも再犯の時期に近接しているし、それはまた一般的に利用されるアウトカムだからである。したがって、これらの基準に合致する効果量によって、異なった研究同士のプログラムの効果についての比較が可能になるのである。そのような効果量が算出できないときは、この基準に最も近く合致するような効果量を使った。たとえば、財産犯は粗暴犯より一般的であり、再収容よりは再起訴に基づいた効果量がより望ましく、12か月にできるだけ近い期間サンプルを追跡した効果量がそれ以外のものよりも望ましい。それぞれの独立した評価研究によって、この「一般的再犯」データセットに対して唯一無二の効果量が算出された。この一般的再犯データセットは、それに続く分析におけるメイン・データセットとして利用した。

われわれはまた、さらに4つのより具体的なデータセットを作った。再逮捕、再収容、そして薬物リラプスのアウトカムに対するデータセットである。これらのデータセットを作るに当たり、そこからできるだけ少数の効果量を選ぶようにした。これらのデータセットのいずれかに対して複数の効果量が得られるときは、以下のような効果量を優先した。すなわち、(1) 一般的である（例：特定

の罪種だけでなく、全罪種をカバーしているようなもの)、(2) 二値変数である、(3) サンプル・メンバーを12か月追跡調査している。もし、ある研究がこれらのアウトカムのどれかを報告していないときは、その研究はその特定のデータセットに寄与できないこととなる。たとえば、もしある研究が逮捕アウトカムのみを報告している場合は、この研究は逮捕データセットには寄与できるが、再収容や薬物リラプスのデータセットには寄与できない。

3.5 研究コーディング・カテゴリーの詳細

このレビューで採用したコーディング用紙は、付録-2のとおりである。これらのコーディング用紙は、階層的な構成となっており、研究内の効果量がネストされていることが明確にわかるようになっている。これらの用紙を使えば、それぞれの評価研究から効果量の数がいくつであれコードすることができる（この問題に関する議論については、Lipsey and Wilson (2001) を参照）。

コーディング用紙は、治療の特性、研究参加者、研究の方法、アウトカム指標、および観察された効果の方向と大きさについての主要な性質を取り出すものである。2人の評定者が、各研究を査定した。評定者間の食い違いは、本研究代表者の1人が調整した。

3.6 統計的方法と取り決め

各比較評価ごとに効果量を計算した。二値変数のアウトカム測定値には、オッズ比効果量が最適であるので、それを二値変数のアウトカムとして使用した (Lipsey and Wilson, 2001)。連続尺度における犯罪行動の指標は、標準化された平均差効果量 (standardized mean difference effect size) を用いてコーディングした。これらの効果量は、治療群が対照群よりも望ましい結果を出した場合（たとえば、再犯や薬物使用の低下など）、正の値を取る効果量で表すようにコードされる。オッズ比効果量 (ES_{or}) は、次のように定義される。

$$ES_{or} = \frac{P_c / (1 - P_c)}{P_t / (1 - P_t)}$$

ここで P_c は対照群において、あるイベント（たとえば、再逮捕）が生じる確率であり、 P_t は治療群において、同一のイベントが生じる確率である²。標準化された平均差効果量 (ES_d) は次のように定義される。

$$ES_d = \frac{\bar{X}_c - \bar{X}_t}{S_{pooled}}$$

\bar{X}_c は対照群の平均値であり、 \bar{X}_t は治療群の平均値、 S_{pooled} は、プールされた群内標準偏差あり、それは次のように求められる。

$$S_{pooled} = \sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{(n_t - 1) + (n_c - 1)}}$$

Sは治療群の分散、 s_c^2 は対照群の分散、 n_t は治療群のサンプル数、 n_c は対照群のサンプル数である。オッズ比効果値と標準化された平均差効果量は、Hasselblad and Hedges (1995) によって開発された方法を用いて統合された。具体的には、平均差効果量がオッズ比効果量尺度上に変換された。これらの効果量についてのわれわれの分析は、Lipsey and Wilson (2001) とWang and Bushman (1999) による統計的アプローチを利用したものである。具体的に言えば、われわれは分散の逆数を使った方法を用い、真の治療効果は、研究間の測定された相違（たとえば、コード化された研究の特質）と、測定されていない相違の双方の関数として変動すると仮定している。研究間の測定されていない相違を把握するために、次のように、各効果に対して計算した母数効果の重みづけにランダム効果因子を加算した。

$$V^* = V + V_\theta$$

² われわれは、オッズ比の逆数を用いていることに留意されたい。というのは、対照群に比べて治療群の再犯可能性が低いときに、1より大きな値となるようにしたいからである。

v は標本誤差の分散， v_0 は効果量の分布から推測されるランダム効果分散である。

また，われわれの分析においては，D. B. Wilson が作成したStataマクロ・プログラムも用いた³。これらのマクロ・プログラムは，上述のランダム効果分散を計算し，統合平均効果及び効果統計量の均一性などのさまざまな統計的数値を計算するものである。さらに，われわれはこのマクロ・プログラムを使って，混合効果モデルが最尤度によって推測されていると仮定し(Raudenbush, 1994; Overton, 1998)，分散分析と回帰のメタ・アナリシスのアナログ研究を通して，研究特性のうちどの特性が観察された効果と関連しているかを決定した(Raudenbush, 1994; Overton, 1998)。公表バイアス分析には，「メタ・バイアス」(公表バイアスのために2つの検定を行うもの)と「メタ・トリム」(公表バイアスのための統計的相関を計算するもの)，Thomas J. Steichen が作成した Stata マクロ・プログラム (これらマクロの双方は，Stata の「ネット・インストール」コマンドによって利用可能である)。そして最後に，Hedges and Pigott (2001)による方法を利用して検出力検定を行った。

3.7 質的データの扱い

この系統的レビューに，質的研究を含める予定はない。しかし，本レビューの今後のアップデートにおいて，そのような手法を用いている専門家からの提言や，彼らとの協働についてはやぶさかではない。

4. 結果

4.1 適格研究の概要

われわれの最初の検索によって，233 の適格研究の候補が見出された。われわれは，これらの研究のうち，229件について論文を入手した。こうして検索さ

³ 現時点では，David Wilsonがこれらのマクロ・プログラムを以下のサイトで公開している。
<http://mason.gmu.edu/~dwilsonb/ma.html>

れた研究のうち、われわれの選択基準に合致したのは53の研究であった。これら53の研究においては、1つの研究が複数の評価を行っている場合があったので、66の独立した評価が報告されていた。具体的には、9つの研究が複数の評価を報告していた。これらのうち7つの研究は2つの評価の報告をしており、1つの研究が3つの評価を、もう1つの研究 (Tunis et al., 1995) は、5つの独立した評価を報告していた。53の研究からコードされた66の独立した評価が、最初のデータセットであった。

アップデート版の検索では、114の研究が候補となった。これらのうち13の研究が選択基準に合致すると決定された。しかし、これらのうちの5つの研究が、同じサンプルを用いて別の評価を行っていたので、8つの研究が独立した研究として残った。これらの8研究は、それぞれ1つのプログラムを評価したものであったので、それぞれの研究から1つずつの効果量が最初のデータセットに加えられた。結果、このアップデート版のデータセットにおける独立した効果量の総数は、74となる。

対象となった評価研究のうち、圧倒的大多数がアメリカにおいて実施されたものである。すなわち、これらの評価研究のうちの65 (88%) がアメリカにおいてなされ、4つの評価 (5%) がカナダで行われ、3つ (4%) がオーストラリアにおいて、そして1つ (1%) がイギリスで、1つ (1%) が台湾において実施された。およそ半分(38, 51%)の評価研究が、学術雑誌の論文、または書籍として発刊されたものからコードされたものであり、他の34の評価研究は公表されていないテクニカル・レポートや政府刊行物からコードされた。発刊の時期に関しては、およそ60%の評価研究 (42)が1999年から後に発刊されたものである。興味深いことに、評価研究の4分の3は1996年以降に発刊されたものであるが、Pearson and Lipton (1999) のレビューに含まれているもののうち最新のものは 1996年発刊である。つまり、本レビューに含まれている評価の大多数は、彼らの行った以前のレビューには含まれていないということになる。

4.2 研究間の統合平均効果

74の評価研究のうち73が、少なくとも1つの釈放後再犯の指標を報告している(残り1つの評価は、薬物使用アウトカムを報告しているのみである)。われわれ

の効果量選択基準を適用すれば（これらの基準については、3.4を参照）、73の一般再犯オッズ比のうち、およそ84%が治療群が対照群に比べて再犯が少なかったことを示しており、オッズ比のおよそ15%が逆の結果を、そしてオッズ比の1%が再犯率が同等であることを示していた。一般再犯オッズ比の分布を見れば、1つの評価は外れ値（OR = 0.016）⁴であることが示唆されるが⁴、この効果量を分析から除外してもしなくても、われわれの報告は適用可能である。

表-1には一般再犯指標のランダム効果平均オッズ比を示している。このアウトカムについての平均オッズ比は1.34で、95%信頼区間は下限が1.21、上限が1.47の範囲であった。このことは、概してこれらの薬物乱用治療プログラムに参加することは、治療後の再犯の低下と関連していることを示している。この効果量をパーセントに変換することによって、一層直観的にこの効果量を把握することができる。さらに進んだ理解のために、対照群の再犯率を35%であると仮定すると（これは元のレビュー中の全対照群の重みづけされていない平均再犯率である）、全般的な平均オッズ比を変換すれば、治療群の再犯率は約29%という値かになる。すなわち、治療を受けることは、再犯率の17%の低減と関連するということになる（すなわち、

$$\frac{29\% - 35\%}{35\%} \approx -17\% \quad)^5$$

もし、負の値を取った外れ値を除外すれば、平均効果量は1.38（95%信頼区間 = 1.26 – 1.51）となる。

一般再犯指標の分布は、標本誤差だけでは説明できないかなり大きな広がりをもっており（ $Q = 706.26$, $df = 73$, $p < 0.001$ ）。このことから、治療プログラムの特質、研究の方法、サンプルの特性が観察された治療効果の大きさに影響を与えていることが示唆される。この可能性については、この後の項で検証していく。

⁴ このオッズ比を対数化オッズ比に換算すると-4.61となり、絶対値が2以上である対数化オッズ比はこのほかには1つもない。このオッズ比は、薬物置換療法プログラムの効果量である（Magura et al., 1993 : 女性サンプル）。

⁵ 仮に対照群の再犯率が50%だとすると、この効果量では治療群の再犯率が43%ということになり、再犯率14%の減少となる。

表-1はまた、4つのアウトカム・データセットのランダム平均オッズ比を示している（このデータセットについては 3.4 を参照）。この表を見ると明らかなのは、アウトカム指標として再起訴を用いて評価すると、平均オッズ比は最大となる。一方、再収容を再犯指標として用いた場合、平均オッズ比は最小である。われわれの使用した一般再犯指標は、最もよく利用されているものであると思われる。というのも、犯罪者について研究をする状況が違えば、それに対応する刑事司法システムの担い手も変わるが、それによって異なった影響を受けることの最も少ない犯罪行動の指標であるからである。さらに、この指標は研究間でのプログラム効果を最も比較しやすい指標であり、そのために、われわれは本項（以下）で報告する分析においてこの指標を用いたのである。

興味深いことに、薬物乱用治療の効果を査定するために、釈放後の薬物使用を使った研究は、74の独立した評価研究のうちわずか22しかなかったのである（図-3参照）。これら22の独立した評価研究のランダム効果平均オッズ比は、1.28（95%信頼区間、0.92-1.75）であった。ここでも対照群のリラプス率を35%と仮定し、この大きさのオッズ比を換算すると、治療群参加者のリラプス率は30%となり、リラプス率において15%の低下である。しかしながら、この平均オッズ比は統計的に有意ではなかった。有意差が出なかったのは、検出力の欠如によるものではない。というのも、われわれが事後検出力分析をしたところ、この分析が小さな効果（すなわち、対数化オッズ比は 0.20）を検出する検出力は 0.30であった⁶。さらに、効果量の分布は標準誤差のみから期待されるものよりずっとばらつきが大きく（ $Q = 197.97$, $df = 19$, $p < 0.001$ ）、ここでもまた、調整変数が、薬物リラプスのオッズ比のばらつきかのいくらかを説明することが示唆される。

4.3 緩衝効果の分析

これまでの分析では、効果量の分布が偶然の要因だけで期待されるものよりもばらつきが大きいことが示された。このことから示唆されるのは、研究の方法、サンプル、介入の違いなどに重要な差異があり、それらが効果量のばらつき

⁶ 検出力検定はすべて、両側5%水準で実施し、観察データを用いたので（すなわち、標準化後誤差、研究数など）、これらは事後検定である。

に何らかの影響を与えているということである。レビューに含まれた個々の研究から得られる情報をコーディングして研究間の重要な相違点をとらえようと試みた。しかし、どれだけ多くの関係する研究特性をコード化できるか否かは、研究論文執筆者が記載している内容の質に左右される。

検証された最初の調整変数は、介入の主な種別である。上述のように（セクション1を参照），コード化された評価研究には主に4つのタイプがあった。すなわち、TC、カウンセリング・プログラム、ブートキャンプ、薬物維持療法である。評価研究の大多数は、TCに関するものであった(35)。もう1つ多数の評価研究があったものは、カウンセリング・プログラムを査定するものである(26)。ブートキャンプと薬物維持療法については、わずか一握りの研究があったのみであり、それぞれ2本、6本であった。さらに、4つの評価研究には、カテゴライズができるだけの詳細な記載がなかった。

表-2を見ると、介入法の違いによって一般的再犯オッズ比に大きなばらつきがあることが分かる ($Q = 12.80, df = 3, p = 0.05$)。平均すれば、TCとカウンセリングによる介入は、一般再犯を有意に減少させている。特に、TCプログラムの評価研究では、平均オッズ比は1.40であった (95%信頼区間, 1.14 - 1.71)。これは、対照群の再犯率をここでも 35%であると仮定するなら、これらのプログラム参加者の再犯率は28%となる。カウンセリング・プログラムの評価研究では、平均オッズ比が1.53 (95%信頼区間, 1.20 - 1.94) であった。この平均オッズ比は、対照群の再犯率を35%であるとするなら、再犯率 26%に当たる。

一方、ブートキャンプと薬物維持療法の平均オッズ比からは、一般的にこれらのプログラムが統計的に有意な再犯低下と関連がないことを示している。より具体的に言えば、薬物事犯者に向けたブートキャンプ・プログラムの評価研究は2本、このレビューに含まれている。その評価研究のいずれにおいてもオッズ比は小さい正の値であり(1.06と1.15)、そのいずれのオッズ比もか統計的には有意でない。この2つのブートキャンプ評価研究の固定効果平均オッズ比は、1.10 (95%信頼区間, 0.48 - 2.50)である。薬物維持療法に関しては、これらのプログラムを検証した評価研究から6つの効果量が求められた。6つのオッズ比のうち3つは1より小さく、その中には前述の外れ値が含まれていた。薬物維持療法を評価したランダム効果平均オッズ比は、0.57 (95%信頼区間, 0.34 - 0.95) で、これはこのプログラムの参加者が、非参加者に比べて、平均して有意に高い再

犯率を有したことを示している。この平均値は、負の値を取る外れ値に大きく影響を受けているが、この外れ値を分布から取り除いても、薬物維持療法プログラムの平均効果量は1.99で、95%信頼区間は0.71 - 1.67であった。つまり、薬物維持療法が再犯率に及ぼす影響の程度には不確定なところがあるが、現在のエビデンスは、これらのプログラムは一般的に再犯を大きく抑制することはないことを明確に示している。さらに、ブートキャンプが薬物事犯者の再犯に及ぼす効果は、非常に小さい。

予備的な緩衝効果の分析を行った結果、オッズ比といくつかの調整変数との関係は、TCまたはカウンセリング・プログラムの評価研究によって異なることが分かった。したがって、一連の並行した分析においてTCとカウンセリング・プログラムについて、別々の緩衝効果の分析を行った。ブートキャンプと薬物維持療法のオッズ比は、この分析からは除外した。

ここで重要なことは、表-3から11に提示した緩衝効果の分析は限定的な検出力しか有していなかったという点である。これらの緩衝効果分析で小さな効果を検出する検出力は、およそ 0.10 から 0.40 のレンジであった⁷。これらの分析は限定的な検出力しか有しておらず、統計的に有意になりやすいのは、大きな効果同士の比較のみであった。言い換えれば、実態的に意義のある効果であっても、これらの分析における検出力の低さゆえに、その多くは典型的な有意水準（たとえば $p < 0.05$ ）において、統計的に有意にはならなかったということである。これらの解析における検出力の低さを解決するために、偶然だけでは10%未満の生起確率しか有さないような相違は（つまり、 $p < 0.10$ ）、統計的に有意であると解釈することとした。

表-3から11に提示された緩衝効果分析におけるもう1つの限界は、これらすべての分析は二値変数であったということである。残念ながら、利用するデータ・セットが少数であるとき、多変量解析を行うことは問題が大きく、そのような分析の結果は小さな変化（たとえば、1つの観察データを削除した）に対して非常に感度が大きくなる。したがって、これら二値変数の知見は 擬似効果

⁷ もし有意水準を0.10まで引き上げるなら、これらの緩衝効果分析で小さな効果をも見出す検出力は、およそ0.20から0.50までのレンジとなる。この検出力は、まだ標準的な0.80よりは低い。したがって、有意水準を0.10としてもなお検出力は限定的なままである。

に脆弱で、その結果、その緩衝効果分析は参考程度にとどめておくべきものとなる。

表-3, 4, 5はTCのみのオッズ比を用いた二変量緩衝効果分析結果を示している。表-3では研究手法別にコードされた一般再犯オッズ比の違いを示している。最初の調整変数、「全般的な方法の質」は、各評価研究の内的妥当性についての4件法の順序尺度の値である。この4件法のカテゴリー分けは、メリーランド大学科学的方法尺度 (Scientific Methods Scale) (Farrington et al., 2002 参照)と同様のものである。方法論的に最も質の低いものは、弱い準実験的デザインであり、これを用いた研究では、介入前に治療群との比較可能性を欠く比較を行うものである。その次のレベルの方法は、「標準的準実験的デザイン」であり、その特徴は、治療群と重要な客観的変数において介入前の時点でわずかに異なる対照群を用いて比較をするものである。「厳密な準実験的デザイン」とは、重要な客観的変数(例: 年齢, 性, 前歴, 薬物使用歴)に関して十分に比較可能である治療群と対照群を用いるもの、または若干相違点のある治療群と対照群を用いて評価をするが、重要な変数についての既存の相違を統制した多変量解析を行うものである。最も質的レベルの高い研究方法は、「実験的デザイン」であるが、それはランダムに被験者を各条件に割り付けるもので、かつ脱落の問題がないものである(付録-2のコーディング用紙を参照) 8。

この基準に基づけば、採用された評価研究の大半は方法論的に弱い。TCの評価研究35のうち、厳密な準実験、または実験的デザインとされたのは13(37%)であった。一番数が多かったのは、「標準的準実験デザイン」であり、このランキングに見合ったのは、35の評価研究のうちの15あった。ここで重要なのは、方法論的に最高レベルであった3つの評価研究の平均オッズ比は、統計的に有意であったということであり、つまり、TCの効果は方法論的に弱い評価研究によってのみ実証されたものではないということである。

8 われわれは2種類の脱落の問題をコード化した。全体的に脱落が20%以上あるとき、または第一著者が脱落者が非脱落者と根本的に異質であったと指摘している場合、全般的脱落と定義した。脱落の偏りも同様に定義した。すなわち、両群の脱落差が20%以上であるとき、または第一著者が、脱落によって治療群と対照群の比較可能性が大幅に損なわれたと指摘しているとき。

方法論の質における各レベルの平均オッズ比は、弱い正のトレンドを示している。つまり、方法論的に一番質の低い評価研究のオッズ比が最小であり、方法的な質が最も高い評価研究のオッズ比が高かったのである。分散分析のメタ・アナリシス的なアナログ研究では、方法論のレベルのバリエーションは、統計的に有意であった ($p = 0.099$)。このことから言えるのは、方法論的に厳密な評価研究ほど治療効果について強力なエビデンスを得ることができたということである。

表-3からわかることはまた、コードされた研究法の特質のほとんどは、TCプログラムの評価研究における治療効果と関連がなかった。すなわち、ランダム付け、被験者レベルのマッチング、多変量解析など、治療群と対照群の間の既存の相違を統制するための方法論的な手法は、どれも効果量とは関連がなかった。さらに、表-3によれば、ほとんどすべての平均オッズ比は、統計的に有意であり、このことからTCの効果は方法論が異なっても一貫していることが分かった。

緩衝効果分析によれば、出版された研究は、出版されていないものよりも統計的にみてより大きな効果量を示していた。このことは、TCの評価研究における公表バイアスの存在を示している。これらの評価研究における公表バイアスをさらに検証するため、公表バイアスのための統計的検証を実施した。具体的には、**Begg and Mazumdar (1994)**と **Egger, Smith, Schneider, and Minder (1997)**のテストの両方を、公表バイアス検証のために実施した。そして、統計的により強力な**Egger et al.**の方法を実施し、公表バイアスのエビデンスを見出した。すなわち、標準化された効果の予測値の回帰の精度に対する切片はゼロに等しいという帰無仮説は却下された ($p = 0.001$)。この結果を受けて、公表バイアスを説明するため、**Duvall and Tweedie (1997)**の“trim and fill”法を実施した。この方法によって、14つの効果量を元の分布に加えると、その結果、平均ランダムオッズ比が1.05、95%信頼区間 .94-1.17 にまで下がった($Q = 415.97$, $df = 48$, $p < 0.001$)。ここでも対照群の再犯率を35%であると仮定し、平均オッズ比に修正された公表バイアスを治療群の再犯率に変換すると、およそ31%となる。したがって、上で述べた結論は、trim-and-fillモデルでは頑健なものではなくなる。しかし、ここで取り扱っているケースのように研究間に多様性が大きいときは、このモデルは過剰に補正する傾向があるとも言われている。

表-4は、一般再犯オッズ比とサンプル特性の間の同様の二変量緩衝効果分析結果を示している。研究者によって一貫して報告されているサンプル特性は4つである。すなわち、年齢(少年または成人)、サンプルの性別構成比、人種構成比、犯罪種別(粗暴犯か非粗暴犯か)である。われわれの分析では、これらのどのサンプル特性も効果量とは統計的な、または大きな関係は見出せなかった。ここでもまた、注目すべきことは、表-4に示したほとんどすべての平均オッズ比は、統計的に有意であり、少なくともある程度は大きな値であった。すなわち、TCは多くの異なったタイプのサンプルに対しても効果的であったということである。

各介入法の特質もまたコード化した。これらの調整変数としての特性を分析した二変量分析を表-5に示した。6種の介入法の特質をコード化した。すなわち、義務的アフターケア、介入場所(ジェイルか刑務所かなど)、治療期間、プログラムの成熟度、参加形態の違い(完全に自発的か、少なくともある程度の強制があるのかなど)、プログラムの参加人員・平均参加者数(これは表-5には記載されていない)⁹。ここでもまた、コードされた特性はいずれも、効果量とは統計的に有意ではなく、大きな関連も見出せなかった。そして、TCの効果のエビデンスは、コードされた治療特性のどのバリエーションにおいても確固としたものであった。具体的に言えば、アフターケア治療が義務化されたTCプログラムの評価、または全受講者が自発的参加であるプログラムの評価では、他のプログラムよりもいくらかの大きな効果量が算出されているが、これらの治療形態の特性は効果量の大きさとは関連がなかった。同様に、治療期間の短いプログラムは長いプログラムよりもやや効果が小さかったが、この差異は有意なものではなかった。実際、これらの評価研究では、治療法の特質についてコード化された差異に関わらず、TCプログラムの効果についてのエビデンスを見出した。このように、ここでもまた、エビデンスの示唆するところは、何点かの主要な評価特性に関わらず、TCプログラムの参加者は非参加者に比べて再犯率が低かったということである。

⁹ (対数化) オッズ比とプログラムの許容人員・平均参加者数は関係は、表-5には記載していない。というのは、これらの関係は、メタ回帰を用いて検証した場合、表-5のフォーマットには適さなかったからである。プログラム許容人員の非標準化係数は、-0.0000021であまりに小さすぎ、統計的にも有意ではなかった ($p = 0.75$)。

カウンセリング・プログラムに対しても、同じような一連の分析が行われた(表-6, 7, 8参照)。ここでもほとんどの評価研究には、方法論的な問題があった。ほとんど4分の3を超える研究(73%)が「弱い」準実験デザインであるか、「標準」準実験的デザインであると評価された。犯罪者を治療条件にランダムに割付けをした実験的デザインを採用している研究は、わずか2つしかなかった。このような方法論的厳密性を欠くことは、特に問題である。というのも、方法論的厳密度の高い評価研究は、一番低いものよりも平均オッズ比が小さい値を取り、有意ではなくなるからである。具体的には、「厳密な」準実験的デザイン、または「実験的デザイン」の平均効果量は、それぞれ 1.33 (95%信頼区間, 0.87-2.05), 1.09 (95%信頼区間, 0.52-2.29)であり、これらのいずれも統計的には有意でなかった。方法論的厳密さのレベルによってオッズ比を比較した統計的検証の結果、有意差はなかったということから、再犯減少におけるカウンセリング・プログラムの有効性の最も有力なエビデンスは、方法論的に弱い評価研究によるものであるということが言える。

オッズ比の大きさと統計的に有意な関連のあった唯一の方法論的変数は、脱落の偏りと多変量データ解析であった(表-6参照)。特に、多変量データ解析を採用した評価研究は、それ以外の評価研究よりも効果量が大きかった。また、脱落の偏りが明白であった評価研究は、脱落に偏りが無い評価研究よりもオッズ比が小さかった。さらに、大きな全般的脱落のない評価研究は、統計的には有意でなかったが、他の評価研究よりも大きな平均効果量を産出していた。またほかにも注意すべき点としては、公表されている研究と未公表の研究を区別する調整変数によるカウンセリング・プログラムの効果の大きさに差はなかった(表-6)。この結果は、公表バイアスの他の検定結果と一致する(この表には不記載)。すなわち、Begg and MazumdarとEgger et al.の公表バイアスの検定は、帰無仮説を棄却しなかった(公表バイアスはないということ)。

サンプル間の相違を説明する4つの調整変数のうち2つが、効果量と統計的関連があった(表-7)。成人サンプルを用いた評価は、少年サンプルを用いた評価よりも平均オッズ比が統計的に大きかった。同様に、女性のサンプルを使った評価研究は、男性または男女混合のサンプルよりも統計的に大きな平均オッズ比を示していた。実際、事後評価では、この3つ全部の平均オッズ比は、統計的にそれぞれ相違していた。サンプルの人種構成は、効果量には何の統計的な大

きな関係もなかった。実際、カウンセリング・プログラムは、すべての人種において再犯減少に効果があった。治療特性に関しては、確立されたカウンセリング・プログラムと自発的プログラムが、他の評価研究よりも統計的に大きな効果量を示していた（表-8参照）。このほかのコードされた治療特性は、効果量と主要な、あるいは統計的な関連を示していない。TCの評価研究の分析と一致する点であるが、義務的アフターケアのあるプログラムは、アフターケアのないものよりも大きな平均効果量を示していたが、その差異は有意ではなかった。ここでも注意すべき重要な点は、調整変数分析は検出力が弱く、それゆえにこれらの所見も（さらにその他の有意ではない所見も）、検出力が弱いゆえのことかもしれないということである。

最後に、リラプスのオッズ比のバリエーションを予測する調整変数の影響を調べた（表-9, 10, 11を参照）。これらの分析に用いられた効果量の数（22）が限定されているため、一次的治療の種類ごとに別個の分析をすることができず、それゆえに、これらの分析においては、一次的治療の全種類を同時に分析した。

おそらく最も重要なことは、薬物リラプス防止の効果量は、治療プログラムの種類によって違っていたということである（表11）。興味深いことに、薬物維持療法プログラムは、最大の平均効果量（2.10）を示していた。治療共同体は、その次に大きな効果量であり（1.33）、それは治療共同体の一般再犯の平均効果量（1.40）と類似したものだった。しかし、一般再犯平均効果量とは異なり、薬物リラプス効果量は、統計的に有意ではなかった（ $p=0.13$ ）。これは、大部分は効果量の数が比較的少なかった（すなわち統計的パワーが小さかった）ことによる。ブートキャンプとカウンセリング・プログラムはいずれも、1より小さい平均効果量であり、これは負の平均治療効果を示している。これらの結果が示すことは、薬物維持療法プログラムは薬物リラプス防止に平均して大きな効果があり、治療共同体は同じアウトカムに対して中程度の効果があるが、カウンセリングとブートキャンプには、一般的に薬物リラプスを抑制する効果がないということである。

薬物リラプスの調整変数分析からわかった興味深いことは、統計的に有意な平均効果量はほとんどなかったということである。この結果が示すことは、わずかな例外を除いて、方法論、サンプル、治療特性の違いにかかわらず、拘禁

下の薬物乱用治療は、一般的に釈放後の薬物使用を抑制しないということである。治療プログラムの違いを超えて、以下のような特質を有する評価研究だけが統計的に有意な($\alpha = 0.10$) 1より大きい平均オッズ比を示した(つまり、薬物使用を統計的に有意に低下させた)。すなわち、ランダム割り付けをしている、参加者のマッチングをしていない、成人サンプルを用いている、女性サンプルを用いている、義務的アフターケアを有している、自発的な参加を必要としている。最後に、公表された評価研究は公表されていないものより、いくらか大きな効果量を示していたものの、公表バイアスの統計的テストは、これらの評価における公表バイアスの存在を示していなかったことも特筆すべきである。

5 結論

全般的に言って、拘禁下における薬物乱用治療の評価研究に対するこのメタ・アナリシスによって、このようなプログラムは再犯防止に対してある程度の効果があることがわかった。一般再犯オッズ比の84%が、対照群よりも治療群の方が勝っていた。さらに、ランダム効果平均オッズ比は、1.34であり、これは対照群の再犯率が35%であるとすれば、治療群の再犯率を29%と換算でき、再犯率の17%の抑制である。しかし、治療プログラムの効果は、治療の特性によって明確に異なっていた。

これまでにあったレビュー同様に(たとえば、Wilson, MacKenzie, and Mitchell, 2005; Pearson and Lipton, 1999), ブートキャンプへの参加が再犯や薬物使用を防止するというエビデンスは見出せなかった。薬物事犯者に対するブートキャンプ・プログラムの独立した評価研究の数は少ないが、われわれの研究も他の研究も一貫して同様な結論であることを考えると、ブートキャンプが一般的に再犯を減少させるとは考えにくい。

同様に、拘禁下における薬物維持療法についても、再犯防止、再使用防止への効果に関するエビデンスは限定的なものであった。見出された7つの評価研のうち3つまでにおいて、治療を受けた者の再犯率が、受けてない者の再犯率よりもいくぶん高かったが、薬物維持療法プログラムが再犯率に及ぼす効果はかなり小さいもので統計的に有意ではなかった。これに反して、薬物維持療法プログ

ラムの全評価研究において、治療を受けた者は、そうでない者よりも釈放後の薬物使用率がいくらか減少している。このように、拘禁下における薬物維持療法は、薬物使用は防止するかもしれないが、再犯の防止はできないであろうと言える。しかし、検出力が限定的であるため、この点については確固とした結論を導くことができない。特筆すべきことは、拘禁下における薬物維持療法の効果に関するわれわれの知見は、コミュニティにおける薬物維持療法に関するより大規模なレビューの結果とは異なっているということだ (Egli, Pina, Christensen, Aebi, and Killias, 2009参照)。これらのプログラムの効果について引続き研究を続けることで、われわれの知識をより一層広げることには有意義な貢献ができるだろう。

治療効果について、最も一貫したエビデンスがあるのは、TCプログラムである。この種のプログラムでは、一貫して釈放後の再犯と薬物使用の防止が示されている。この結果は、方法論を変えても不変であった。事実、最も厳密な評価研究においても、TCプログラムへの参加は、一貫して再犯防止と関連があった。いくつかの異なったタイプのサンプルに対しても、TCの効果が認められた (たとえば、女性のみサンプル、男性のみサンプル、成人のサンプル)。このことから示唆されるのは、TCはさまざまなタイプの犯罪者に幅広く適用可能だということである。しかし、入手できたTCの評価研究には公表バイアスの可能性があるため、われわれの知見はより控え目なものになる。つまり、この分野の研究には公表バイアスがあるとのエビデンスがあるため、TCプログラムの効果は明らかに過大評価されている。TCプログラムの薬物リラプスに対する効果に関しては、再犯率に対する効果と同等の大きさであった。しかし、TCプログラムの薬物リラプスに対する効果を検証した研究の数が少ないことが大きな理由であるが、その効果は統計的には有意でなかった。

カウンセリング・プログラムに関するエビデンスは、これらのプログラムは再犯防止には効果的であるが、薬物使用防止には効果がないということを示すものであった。カウンセリング・プログラムは、成人、または女性の犯罪者を対象としたときに最も効果が大きいようだ。エビデンスによればまた、厳格に自発的に行われたカウンセリング・プログラムは、それ以外のカウンセリング・プログラムよりも再犯防止に効果的である。しかし、カウンセリング・プログラムについての最大のエビデンスは、方法論的に難点のある評価研究によるもの

であったので、われわれの知見も控え目なものとなる。さらに、カウンセリング・プログラムの薬物使用防止に関する効果を評価した研究はわずかしがなく、これらの現存する研究では、カウンセリング・プログラムが薬物使用を防止するかどうか見出すことができなかった。

興味深いことは、すべての緩衝効果分析において、釈放後に義務的アフターケアのある治療プログラムは、そうでないプログラムよりも大きな効果量を産出しているということである。薬物使用効果量以外のすべての分析において、これらの相違は有意ではなかった。しかし、これらの分析の検出力の不足、ただ1つの評価研究を削除したり、加えたりすることへの感度、そしてアフターケアがこの種の介入の効果を強めるという現存のエビデンスを考慮すると、義務的アフターケアを治療プログラムに組み込むことによって、拘禁下における治療の効果が最も増大しやすいであろう。

本研究からの政策決定者への提言は明確である。拘禁されている物質乱用者に有効な介入を実施したいなら、TCのように物質乱用者の抱える複数の問題に強力に焦点を当てるプログラムが最も成功を収めやすいであろうということである。あまり強力でないプログラムからは、期待できる治療効果も小さい。また、現存の研究では、物質乱用者を対象とした矯正ブート・キャンプは、釈放後の再犯防止においても、薬物使用防止においてもエビデンスがない。したがって、政策決定者は、そのようなプログラムには再犯防止の期待をすべきではない。

また本研究は、今後の研究に対して示唆しているものがあると信じている。特に、現存する研究が、あるプログラムの効果を明確に支持したとしても、その治療プログラムのうちのどの要素が最重要であるのか、あるいはそれら要素同士のどのような組み合わせが一番効果を生むのかという点に関してはまだ分かっていない。さらに、この分野の研究においては一般的な方法論的脆弱性があるため、結果について別の説明が可能になってしまう弱みがある（例:再犯の減少は介入ではなく、他の要因によるのかもしれない）。今後、これらの点について研究がなされれば有益であろう。

6 レビューの更新予定

われわれは、キャンベル共同計画のガイドラインに沿って、3年ごとにこの系統的レビューを更新することとしている

7 謝辞

ノルディック・キャンベル・センター、および Jerry Lee Foundation から一部助成に対して謝意を表したい。また、対象とした評価研究の多くのコーディングを手がけてくれた Matthew Makarios にも感謝する。

8 利害対立に関するステートメント

いずれの著者も、既存の、又は計画中の拘禁下の薬物乱用治療プログラムや、薬物乱用犯罪者に対する他の競合する介入法について、いかなる経済的関心も有していない。

9 参考文献

10 系統的レビューに含まれた研究

11 表

表-1 再犯種別平均ランダム効果オッズ比

アウトカム	平均効果量	95% 信頼区間		Q	k ^a
		下 限	上 限		
一般的再犯	1.34*	1.21	1.47	706.26*	73
再逮捕	1.40*	1.27	1.55	204.58*	38
再起訴	1.43*	1.29	1.59	27.64	19
再収容	1.24*	1.10	1.41	319.53*	41
薬物再使用	1.28	0.94	1.75	205.10*	22

a オッズ比の数

* $p < 0.05$, # $p < 0.10$

表-2 治療特性別一般再犯オッズ比

プログラムの種類	平均 ES	95% 信頼区間		k ^a
		下限	上限	
治療共同体	1.40*	1.14	1.71	35
カウンセリング	1.53*	1.20	1.94	26
薬物維持療法	0.57	0.34	0.95	6
ブートキャンプ	1.10	0.48	2.50	2

a オッズ比の数

* $p < 0.05$, # $p < 0.10$

Between $Q = 12.80$, $df = 3$, $p = 0.005$

表-3 方法別一般再犯オッズ比：TCのみ

変数	平均 ES	95% 信頼区間		k^a
		下限	上限	
全般的な方法の質 ⁺				
弱い準実験的デザイン	1.00	0.79	1.25	7
標準的準実験的デザイン	1.44*	1.22	1.69	15
厳密な準実験的デザイン	1.33*	1.09	1.63	11
実験的デザイン	1.90*	1.26	2.87	2
ランダム割付け				
なし	1.31*	1.15	1.48	33
あり	1.90*	1.18	3.05	2
被験者レベル・マッチングの有無				
なし	1.41*	1.22	1.61	28
あり	1.12	0.85	1.46	7
多変量解析				
なし	1.24*	1.03	1.48	15
あり	1.45*	1.22	1.72	20
明確な全般的脱落				
なし	1.32*	1.15	1.53	27
あり	1.32#	0.95	1.83	6
明確な脱落の偏り				
なし	1.37*	1.19	1.58	27
あり	1.17	0.82	1.65	5
公表 ⁺				
なし	1.14#	0.99	1.32	19
あり	1.62*	1.37	1.91	16

a オッズ比の数

* $p < 0.05$, # $p < 0.10$

+ 平均値の差が 5%水準において統計的に有意

+ 平均値の差が 10%水準において統計的に有意

表-4 サンプル特性別一般再犯オッズ比：TCのみ

変数	平均 ES	95% 信頼区間		k ^a
		下限	上限	
年齢				
成人	1.38*	1.20	1.59	30
少年	1.21	0.84	1.76	4
サンプルの性別構成				
女性のみ	1.65*	1.14	2.39	6
男女混合	1.23	0.84	1.79	4
男性のみ	1.36*	1.13	1.64	18
サンプルの人種別構成				
非白人が 50%以下	1.62*	1.21	2.16	8
非白人が 51%以上 70%未満	1.32*	1.03	1.68	12
非白人が 70%以上	1.23	0.93	1.63	8
犯罪種別				
非粗暴犯	1.49*	1.24	1.79	15
混合 (粗暴犯・非粗暴犯)	1.28*	1.02	1.62	9

a オッズ比の数

* $p < 0.05$

$p < 0.10$

† 平均値の差が 5%水準において統計的に有意

+ 平均値の差が 10%水準において統計的に有意

表－5 治療特性別一般再犯オッズ比：TCのみ

変数	平均 ES	95% 信頼区間		<i>k</i> ^a
		下限	上限	
義務的アフターケア				
なし	1.33*	1.11	1.58	18
あり	1.41*	1.12	1.79	10
治療場所				
刑務所	1.33*	1.17	1.52	32
ジェイル	1.56#	0.95	2.56	3
プログラムの成熟度				
新規プログラム（1年未満）	1.33*	1.10	1.59	16
発展途上（1－3年）	1.18	0.80	1.75	4
確立プログラム（3年超）	1.45*	1.16	1.80	13
短期治療（90日未満）				
非該当	1.45*	1.26	1.68	22
該当	1.15	0.79	1.67	3
厳密な自発的治療				
非該当	1.27*	1.08	1.50	11
該当	1.58*	1.37	1.82	17

a オッズ比の数

* $p < 0.05$

$p < 0.10$

† 平均値の差が 5%水準において統計的に有意

+ 平均値の差が 10%水準において統計的に有意

表－6 方法別一般再犯オッズ比：カウンセリング

変 数	平均 ES	95% 信頼区間		<i>k</i> ^a
		下限	上限	
全般的な手法の質				
準実験的デザイン	1.82*	1.25	2.65	8
標準的準実験的デザイン	1.52*	1.12	2.07	11
厳格な準実験的デザイン	1.33	0.87	2.05	5
実験的デザイン	1.09	0.52	2.29	2
ランダム割付け				
なし	1.56*	1.26	1.92	24
あり	1.09	0.52	2.31	2
被験者レベル・マッチングの有無				
なし	1.49*	1.17	1.89	20
あり	1.59#	0.97	2.60	5
多変量解析 ⁺				
なし	1.18	0.86	1.63	9
あり	1.74*	1.37	2.21	17
明確な全般的脱落				
なし	1.64*	1.30	2.07	19
あり	1.15	0.77	1.71	6
明確な脱落の偏り ⁺				
なし	1.69*	1.39	2.05	21
あり	0.71	0.43	1.15	3
公表				
なし	1.56*	1.11	2.19	10
あり	1.49*	1.15	1.93	16

a オッズ比の数

* $p < 0.05$

$p < 0.10$

† 平均値の差が 5%水準において統計的に有意

+ 平均値の差が 10%水準において統計的に有意

表ー 7 サンプル特性別一般再犯オッズ比：カウンセリング

変 数	平均 ES	95% 信頼区間		<i>k</i> ^a
		下限	上限	
年齢				
成人	1.54*	1.41	1.67	20
少年	1.16	0.92	1.46	3
サンプルの性別構成				
女性のみ	2.94*	1.74	4.97	3
男女混合	1.01	0.69	1.48	6
男性のみ	1.67*	1.26	2.21	11
サンプルの人種別構成				
非白人が 50%以下	1.48*	1.19	1.85	5
非白人が 51%以上 70%未満	1.46*	1.21	1.71	7
非白人が 70%以上	1.50*	1.34	1.68	2
犯罪種別				
非粗暴犯	1.48*	1.18	1.86	11
混合（粗暴犯・非粗暴犯）	1.26#	0.98	1.62	12

a オッズ比の数

* $p < 0.05$

$p < 0.10$

† 平均値の差が 5%水準において統計的に有意

+ 平均値の差が 10%水準において統計的に有意

表－８ 治療特性別一般再犯オッズ比：カウンセリング

変 数	平均 ES	95% 信頼区間		<i>k</i> ^a
		下限	上限	
義務的アフターケア				
なし	1.47*	1.16	1.85	21
あり	1.82*	1.10	3.02	4
治療場所				
刑務所	1.58*	1.22	2.04	17
ジェイル	1.42#	1.00	2.00	9
プログラムの成熟度 ⁺				
新規プログラム（1年未満）	1.08	0.82	1.41	8
発展途上（1－3年）	1.43	0.70	2.92	2
確立プログラム（3年超）	1.79*	1.36	2.37	9
短期治療（90日未満）				
非該当	1.44#	0.98	2.12	10
該当	1.58*	1.12	2.23	11
厳密な自発的治療 ⁺				
非該当	1.07	0.60	1.92	4
該当	1.75*	1.26	2.44	14

a オッズ比の数

* $p < 0.05$

$p < 0.10$

† 平均値の差が 5%水準において統計的に有意

+ 平均値の差が 10%水準において統計的に有意

表－9 方法別薬物リラプスオッズ比

変 数	平均 ES	95% 信頼区間		k^a
		下限	上限	
全般的な手法の質 [†]				
脆弱な準実験的デザイン	0.69	0.40	1.21	5
標準的準実験的デザイン	1.46	0.82	2.61	5
厳格な準実験的デザイン	1.21	0.82	1.80	9
実験的デザイン	3.69*	1.50	9.06	2
ランダム割付け				
なし	1.15	0.82	1.59	19
あり	2.58*	1.13	5.92	3
被験者レベル・マッチングの有無				
なし	1.39 [#]	0.97	1.98	19
あり	0.93	0.44	1.98	3
多変量解析				
なし	1.13	0.68	1.89	9
あり	1.41	0.92	2.16	13
明確な全般的脱落				
なし	1.31	0.82	2.07	11
あり	1.18	0.73	1.91	10
明確な脱落の偏り				
なし	1.32	0.87	2.01	13
あり	1.12	0.65	1.94	8
公表				
なし	1.16	0.72	1.87	10
あり	1.42	0.90	2.23	12

a オッズ比の数

* $p < 0.05$

[#] $p < 0.10$

† 平均値の差が 5%水準において統計的に有意

+ 平均値の差が 10%水準において統計的に有意

表-10 サンプル特性別薬物リラプスオッズ比

変数	平均 ES	95% 信頼区間		k^a
		下限	上限	
年齢 [†]				
成人	1.61*	1.24	2.09	16
少年	0.78	0.48	1.24	4
サンプルの性別構成				
女性のみ	2.06#	0.97	4.38	5
男女混合	0.81	0.29	2.26	2
男性のみ	1.18	0.77	1.82	12
サンプルの人種別構成				
非白人が 50%以下	1.85	0.70	4.89	2
非白人が 51%以上 70%未満	1.50	0.88	2.54	7
非白人が 70%以上	1.21	0.73	2.00	7
犯罪種別				
非粗暴犯	1.41	0.80	2.49	8
混合 (粗暴犯・非粗暴犯)	0.94	0.57	1.55	8

a オッズ比の数

* $p < 0.05$

$p < 0.10$

† 平均値の差が 5%水準において統計的に有意

+ 平均値の差が 10%水準において統計的に有意

表-11 治療特性別薬物リラプスオッズ比

変数	平均 ES	95% 信頼区間		k ^a
		下限	上限	
プログラムの種類				
治療共同体	1.33	0.92	1.93	13
カウンセリング	0.77	0.35	1.70	3
ブートキャンプ	0.56	0.16	2.01	1
薬物維持療法	2.10	1.03	4.27	5
義務的アフターケア ⁺				
なし	0.85	0.51	1.43	8
あり	1.79*	1.14	2.82	11
治療場所				
刑務所	1.25	0.89	1.76	20
ジェイル	1.97	0.53	7.39	2
プログラムの成熟度				
新規プログラム (1年未満)	1.12	0.64	1.94	8
発展途上 (1-3年)	1.69	0.78	3.63	5
確立プログラム (3年超)	1.14	0.60	2.18	6
短期治療 (90日未満)				
非該当	1.10	0.71	1.69	13
該当	2.04	0.85	4.93	4
厳密な自発的治療				
非該当	0.93	0.45	1.93	4
該当	1.64*	1.06	2.52	14

a オッズ比の数

* $p < 0.05$

$p < 0.10$

+ 平均値の差が 5%水準において統計的に有意

+ 平均値の差が 10%水準において統計的に有意