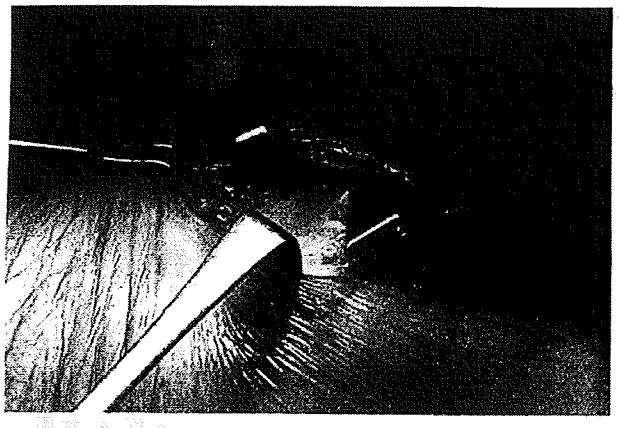


図2 大腿骨前面の骨切り
膝関節伸展位で行うと良好な視野が得られる。

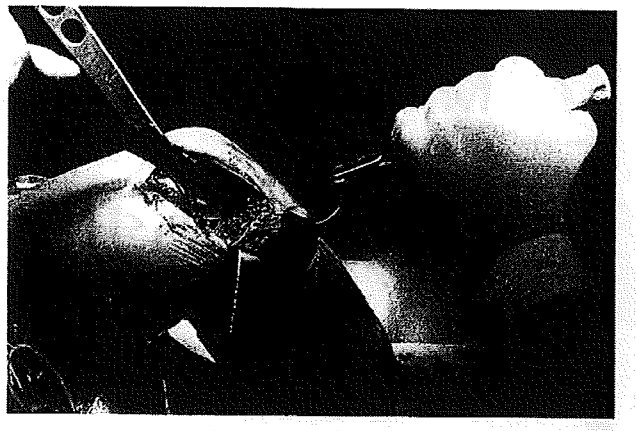


図3 脛骨のステム部やキール部の骨切り
ここまでくると，周囲軟部組織の緊張が低下し，比較的手術が楽になる。

う。骨棘を切除した後，膝蓋骨の lateral facet はやや厚めに残しながら central ridge までを骨切りし，関節内の視野と膝蓋骨の外側への可動性を確保する。術中に牽引操作などにより膝蓋骨骨切り面の不整化を生じる可能性があるので，この時点では膝蓋骨の最終的な骨切りは行わず，視野を確保するための rough cut に留める。膝蓋骨の置換を行わない場合には，rough cut は行わないが，置換する場合には rough cut をこの時点で行っておくと，より良好な視野が得られる。

3)骨切り

従来の TKA では，大腿骨の骨切りと脛骨骨切りを別々のプロセスで行うが，MIS-TKA では術野が小さいため，骨切りの順番にも工夫を要する。

最初に大腿骨遠位の骨切りを側面から専用のガイドを用いて行う。大腿骨遠位骨切り専用ガイドは，通常の内・外側顆に当てる T 字型のガイドと異なり，L 字型で内側顆に当てて骨切りを行う。したがって，内側顆の軟骨や骨欠損が強い症例では，外側顆の高さを考慮して骨切りを行う必要がある。

次いで，大腿骨の残りの骨切りは後回しにして，脛骨近位の骨切りを行う。脛骨の骨切りも専用のガイドを用いて前内方 45° から行う。脛骨外側部の視野が十分に得られない症例では，脛骨の骨片を 2 分割または 3 分割しながら視野を確保し，徐々に切除を進める。外側部の骨切りは視野が悪いため，手術手技に習熟することが大切である。また，骨切除などは伸展位で行うと比較的容易である。脛骨近位の骨切りが終わると，その厚みの分だけ，すなわち extension gap の分だけ術野に余裕ができるため，周囲軟部組織の緊張が低下し，その後は比較的手術が楽になる。

これを利用して，大腿骨前面や後面，チャンファーの骨切りを行う。その際も大腿骨前面の操作は膝関節伸転位で，後面の操作は屈曲位で行うと良好な視野が得られる(図2)。大腿骨の回旋は後顆からの 3° 外旋位と Whiteside line を参考に決定する。さらに，脛骨のステム部やキール部の骨切り(図3)，rough cut した膝蓋骨の最終的な骨切りを行う。

これらの操作が終了すると，膝関節後方の視野もさらに良くなり，後方の軟部組織剝離や骨棘の切除，ある程度の外側の骨棘処理などが可能となるため，この時点でこれらの処置を行う。RA に対する滑膜切除を行う場合は，ここまで骨切りを進めてから行うと，膝蓋上嚢の展開もよく，容易である。

4)インプラントの挿入

骨切りが終了したら通常通りトライアルを挿入し，アラインメントなどを確認してから，インプラントを挿入する。まず，伸展位で脛骨コンポーネントを挿入した後，屈曲位にして打ち込み，セメント固定を行う。次いで，大腿骨コンポーネント，膝蓋骨コンポーネントの順に挿入，セメント固定する。視野が悪い場合には，軟部組織を巻き込まないように十分注意しながらインプラントを挿入することが大切である。

3 後療法

術翌日より大腿四頭筋訓練と CPM 装置を用いた可動域訓練を開始し，術後 3〜4 日で荷重歩行を許可する。平地歩行訓練，階段昇降訓練，希望により自転車訓練を進め，早ければ 10 日前後，通常 2〜3 週で退院とする。

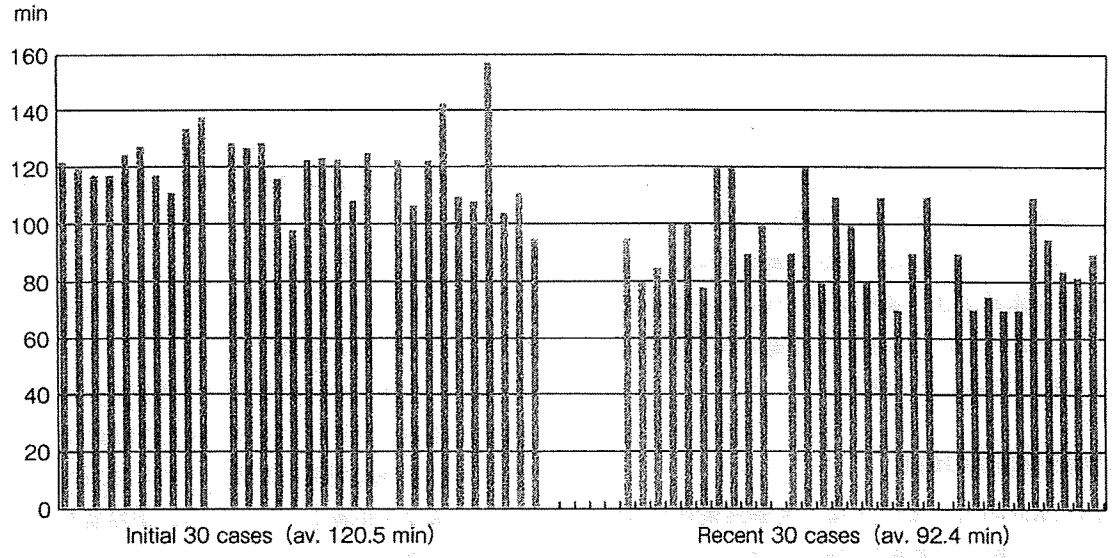


図 4　手術時間の learning curve による変化
初期の 30 例(平均 120.5 分)と 3 年後の 30 例(平均 92.4 分)の比較。手技を習熟すれば，手術時間も徐々に短くなるが，正確な骨切りや安全な手術手技を最優先に行うことが大切である。

4 考察

MIS-TKA の目標は従来の TKA における手術侵襲を最低限に抑えながら，これまでと同等の手術成績を得ることである。大腿四頭筋などに対する侵襲は小さければ小さいほど，術後の疼痛軽減や早期の機能回復には極めて有利である。しかし，術野が著しく制限されるため，それに伴う問題点を 1 つずつ解決していかなければならない。

1)Learning Curve

MIS-TKA は，learning curve が強いことが指摘されている(図 4)。初心者がいきなり MIS-TKA を行うべきではなく，従来の TKA を十分に経験して，膝関節の解剖に熟知し，手術に対する心構えを習得してから行うべきである。また，十分に TKA の経験のある術者も手術器械や手技が大きく異なるため，MIS-TKA の十分な知識とトレーニングを積んでから行うべきである。トレーニングの方法としては，まず MIS-TKA の器械を用いて従来の皮切で手術を行い，器械に十分慣れてから，徐々に皮切や大腿四頭筋への切開を小さくしていく。手技を習熟すれば，手術時間も徐々に短くなるが，正確な骨切りや安全な手術手技を最優先に行うことが大切である。

2)手術適応

MIS-TKA は小皮切から進入し，しかも大腿四頭筋の侵襲も最小限にするため，関節内の視野や操作範囲が限定される。したがって，著しい伸展制限や屈曲制限のある症例，膝蓋骨低位のある症例などでは十分な展開が得られにくいので適応が制限される。また，関節内の広範囲に軟部組織の剝離や骨棘切除を要する症例も適応が難しい。特に外側部の剝離や骨棘切除を行うことは容易ではないため，高度の外側型 OA への適応には限界があると考えられる。さらに，MIS-TKA では関節内の視野を確保するためレトラクターなどによる牽引を多用するため，RA の進行例などで著しい骨萎縮を認める症例では，術中，牽引などにより，骨切り面の不整化を生じやすいので，過度の牽引は禁忌である。

しかし，MIS-TKA のいいところは，手術中に，必要と判断した場合には，いつでも皮切や大腿四頭筋に対する切開を追加できることである。したがって，術中必要があれば，無理をせず，いつでも切開を追加する勇気をもつことが大切である。

3)皮膚切開

一般的な TKA では，傍膝蓋内側皮切が用いられ，その長さは通常短くても 12〜15 cm 程度必要である。皮切の長さは整容的見地ばかりでなく，術後の疼痛や機能回復を考えても可能であれば小さいほうが好ましい。MIS-TKA では，通常 7〜10 cm の皮切で手術が可能である。ただし，皮切長は手術操作そのものよりも，挿入する大腿骨コンポーネントの大きさによって左右される。小柄な女性で，最も小さいサイズの大腿骨コンポーネントを挿入するには約 7 cm の皮切で可能であるが，

平均的な体格の日本人男性では，10 cm 程度の皮切を要する。MIS-TKA では狭い範囲で手術操作を行うため，皮切長の 1 cm の差が術中の操作性に大きく影響する。したがって，術前に大腿骨コンポーネントのサイズを検討して，サイズが大きい場合にはあらかじめ十分な皮切長を確保しておくほうが，手術が容易となる。

いずれにせよ，皮切の長さで最も大切なことは，術中いつでも延長でき，術中に視野が十分得られなかったり，コンポーネントの挿入に困難がある場合には，無理をせず延長することである。

4) 大腿四頭筋の処置

QS-TKA のコンセプトは，大腿四頭筋に侵襲を加えないことである。従来の TKA に伴う大腿四頭筋への侵襲は大きく，これが術後の機能回復訓練のスケジュールに大きく影響する。したがって，この大腿四頭筋への侵襲を最小限に抑えられれば，術後の疼痛を緩和できるばかりでなく，早期の機能回復訓練が可能になる。

しかし，われわれの経験では大腿四頭筋に全く侵襲を加えないで手術が可能であったのは，先に述べたように全症例の 10% であった。内側広筋の大腿直筋への付着はさまざまなタイプがあり，最も末梢まで付着するタイプでは膝蓋骨の中央部付近にまで及ぶ。特に日本人では欧米人に比べ内側広筋が末梢に付着することが多い。このような症例では，最初の切開の段階で，midvastus または subvastus 方向への展開を行う必要がある。われわれは subvastus 方向への展開を行っている。Subvastus は膝蓋骨を反転するためには大きな展開を必要とするが，MIS-TKA のように膝蓋骨を外側にシフトするだけであれば，内側広筋の下縁を 3〜4 cm 展開するだけで，比較的容易に視野が得られる。

5) 関節内の展開

MIS-TKA では小さな皮切から進入し，しかも大腿四頭筋の侵襲を最小限に抑えるため，関節内の展開に工夫を要する。特に骨切り方法とそれに用いるガイドは既存のものをただ小さくして行うだけではなく，根本的な変更を要する。まず，骨切りの順序は，通常は大腿骨，脛骨いずれも，それぞれ十分に展開してから片方ずつ行うが，MIS-TKA では術野を確保しながら骨切りを進めるため工夫を要する。すなわち，術野を確保するための膝蓋骨の rough cut，大腿骨遠位の骨切り，脛骨近位の骨切り，大腿骨の残り 4 面の骨切り，これらの操作によりある程度の視野が確保できてから脛骨のステム部やキール部の骨切り，膝蓋骨の最終的な骨切りの順序で行う。また，後方の骨膜剝離や骨棘の切除，外側の骨棘の処理，RA に対する滑膜切除なども，すべての骨切りが終わってから行うと容易である。また，MIS-TKA ではすべての手技を順番通り行おうとせず，できることから徐々に視野を確保しながら行う工夫も有用である。

6) その他

MIS-TKA では，この他に従来の TKA と異なる工夫がいくつか必要である。

まず，MIS-TKA では膝蓋骨の反転を行わないため，通常の TKA では 90° 以上の屈曲位で行っている手術操作も伸展機構の緊張のために屈曲位では行えず，軽度屈曲位や伸展位で行ったほうが容易なことが多い。特に，脛骨近位の骨切り時に後方や外側部の十分な視野が得られにくいことがあり，その際には伸展位にすると良好な視野が得られる。また，大腿骨前面の骨切りを行う際も，屈曲位では良好な視野が得られにくく，軽度屈曲位で行うと骨切り面を見ながら操作ができるので安全である。いずれにせよ，術中に適宜膝関節を伸展・屈曲しながら手術操作を行うことが大切である。

また，MIS-TKA では視野が小さいため，window technique とよばれる術野を筋鉤やレトラクターなどで動かしながら確保するテクニックが用いられる。すなわち，膝関節内側の手術操作を行う際には内側を牽引して外側を緩め，外側の手術操作を行う際にはその反対を行う。また，先に述べたように術中に膝の屈曲角度をさまざまに変化させながら手術を行う。したがって，助手の役割は極めて大切であり，助手も術前から十分に手術方法を学習し，習熟しておく必要がある。

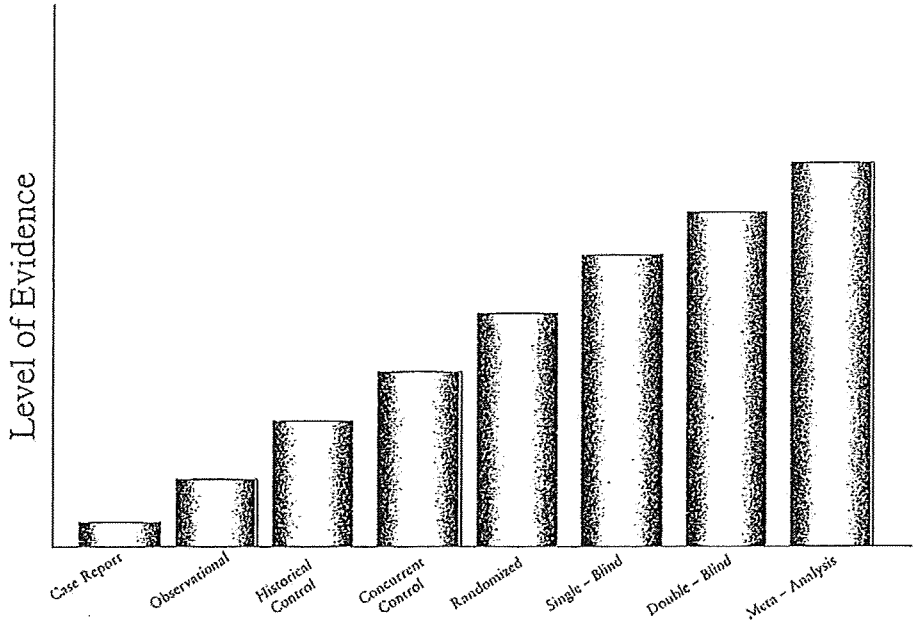
MIS-TKA は従来の TKA に比べ，皮膚切開，大腿四頭筋への侵襲などが明らかに小さい。したがって，これまでと同等の手術成績を得ることができれば，疼痛軽減や早期の機能回復には極めて有利であり，TKA がさらに進歩することは間違いない。今後，手術手技の進歩，instrument のさらなる改良やインプラントそのものの改良などにより，手術成績をさらに向上させることが急務である。

参考文献

1) Bonutti PM, Mont MA, Kester MA : Minimally invasive total knee arthroplasty. A 10-feature evolutionary approach. Orthop Clin North Am 35 : 217-226, 2004
2) Laskin RS, Beksac B, Phongiunakorn A, et al : Minimally invasive total knee replacement through a mini-midvastus incision: an outcome study. Clin Orthop Relat Res 428 : 74-81, 2004
3) Tria AJ Jr, Coon TM : Minimal incision total knee arthro-

plasty : early experience. Clin Orthop Relat Res 416 : 185-190, 2003
4) Tria AJ Jr : Minimally invasive total knee arthroplasty : importance of instrumentation. Orthop Clin North Am 35 : 227-234, 2004
5) 松本秀男，大谷俊郎，松崎健一郎，他：大腿四頭筋温存型人工膝関節―手術技法と問題点．整形外科 57 : 84-88, 2006
6) 松本秀男：特集　最小侵襲の人工膝関節置換術　最小侵襲人工膝関節置換術の手術手技と pitfall．整・災外 49 : 101-107, 2006
7) 松本秀男，他：特集　変形性関節症の保存療法と MIS II　大腿四頭筋温存型 MIS-TKA．関節外科 25 : 525-530, 2006
8) 松本秀男：関節手術の最前線 Part3 最小侵襲人工膝関節置換術．分子リウマチ 3 : 68-71, 2006
9) 松本秀男：MIS-TKA と learning curve．骨・関節・靱帯 19 : 797-804, 2006

Clinical Trials Handbook



Edited By

SHAYNE COX GAD

WILEY

10.8

Musculoskeletal Disorders

Masami Akai

Director, Rehabilitation Hospital,
National Rehabilitation Center Japan, Saitama, Japan

Contents

Clinical Trials Handbook, Edited by Shayne Cox Gad

10.8.1 INTRODUCTION

When we look at past medical history, we can only find a few definite treatments that focus directly on their etiology, such as vaccination for some infectious diseases, vitamin C for scurvy, and, recently, gene therapy. The evidence for presently accepted treatments in our practice is insufficient; it ranges from a scattered presence to total absence. Unfortunately, apart from a few dichotomous results, we cannot expect qualitatively definite results of treatment for all clinical cases. The effect of the interventions is usually assessed quantitatively, comparing before and after treatment, according to certain outcome measures. The key to solving many of the problems in medical practice is to establish valid and reliable methods of evaluating the differences in the outcomes before and after treatment.

I started my postgraduate training as an orthopedic surgeon and then shifted my career to rehabilitation medicine. I am not, by any standard, an expert in all of the subjects covered here: clinical epidemiology, biostatistics, and related fields. However, I have had chances to join teams to organize clinical trials [1, 2]. I would like to describe the necessary steps for conducting a clinical trial from the standpoint of a clinician.

The term *evidence-based medicine* (EBM), which is the integration of best research evidence with clinical expertise and patient values, has already established wide popularity, even in the musculoskeletal field. When we review the clinical trial trends in the musculoskeletal field, we notice several characteristic points, which are obviously different from other areas.

In the musculoskeletal field, there are the typical clinical trials that use placebo-controlled, double-blind methods, such as drug trials for osteoporosis or rheumatoid arthritis. However, if we wish to compare one interventional procedure with another, the double-blind design is often impossible to do methodologically. Only in an open-label trial, where patients know about their treatment, is it possible to conduct the trial under random allocation. With such limitations, we have to suppress various biases as much as possible. Another point that needs attention is the selection of outcome measures suitable for musculoskeletal disorders. Even though there are primary malignant tumors, such as osteosarcoma or a cancer lesion as secondary

metastasis, which are sometimes lethal, the majority of musculoskeletal disorders that develop clinical manifestation are not life-threatening but do impair function. We have to use outcome measures for observing the functional ability of the patients, not for the numerical laboratory data.

In this chapter I would like to discuss the following:

1. Definition of musculoskeletal disorders
2. Background of clinical trials
3. Methodological assessment for clinical trials
4. Preparation of necessary tools for clinical trials
5. Data analysis and interpretation
6. RCT and our future direction

10.8.2 DEFINITION OF MUSCULOSKELETAL DISORDERS

10.8.2.1 Musculoskeletal Disorders

The term *musculoskeletal disorders* refers to conditions that involve the supporting structures of the body, such as the trunk and the bones and joints of the extremities. In other words, musculoskeletal disorders involve locomotor functions that have suffered from trauma, diseases, anomalies, and inevitable aging and are mainly attended to by orthopedic surgeons.

Musculoskeletal disorders are the most common cause of severe, long-lasting pain and physical disability affecting hundreds of millions of people in the world. The extent of such problems and the growing burden is described as "age quake" instead of earthquake. The impact on society from these disorders, combined with the recognition that our health care resources need to be used more efficiently, has led to the organized movement of the "Bone and Joint Decade 2000–2010" [3]. This umbrella organization is attempting to raise social awareness of the suffering and pain of musculoskeletal disorders, as well as the growing burden and cost to society that will come with this age quake. We need to advance clinical research in order to reduce the burden of musculoskeletal diseases.

10.8.2.2 Aims of Treatment

The aims of treatment for musculoskeletal disorders are not to reduce mortality or morbidity but to increase mobility and function, to relieve pain, or to improve quality of life. We have to assess the results of our interventions with a view to the improvement in disability or functional limitation after treatment. In this field, the clinical indicators are neither a 5-year survival rate nor an infantile survival rate. Our aims are to improve the health-related quality of life of the millions of patients suffering from musculoskeletal disorders such as joint diseases, spinal disorders, severe trauma to the body and extremities, and deformity and crippling diseases in children.

When we talk about outcome measures in disability assessment, the key issue to be discussed is what "function" and "disability" mean and how to measure them.

10.8.2.3 Concept of "Construct"

Our function or ability is measured in accordance with conceptual "construct" [4]. We cannot measure this function or ability directly.

In the past, orthopedic surgeons mainly evaluated the amount of disease involvement in view of physiological or biochemical parameters, pathological findings, and impairments such as range of motion or muscle strength. It was thought that these indicators were objective and independent of the unreliable responses or feelings of patients. However, are these indicators truly reflective of the needs of patients and their families to know how much they can expect as a result of treatment? Their expectations and questions would be more like, "Can I walk again?" or "Will she be able to live by herself as before?"

When a certain standardized scale is not available to measure the function or ability of interests, a new construct, according to a concept for measurement, has to be made.

Conventional clinical indicators, such as a 5-year survival rate or an infantile mortality rate, have been recognized as very important health measures in assessing health conditions in society, but lately, the situation has been changing. The recent close-up of patient-based outcome measure is deeply related to this trend [5]. We have to recognize the demand and organize those types of outcome measures for the individual [6].

10.8.2.4 Assessment of Function and Quality of Life (QOL)

Functional assessment, from the physician's point of view, has been mainly used to evaluate problems at the impairment level. Traditionally, orthopedic surgeons have used clinician-based outcomes such as pain score, measurement of range of motion, muscle manual test, or X-ray findings. However, due to various biases and poor judgments of accuracy levels, clinician-based outcomes often affect the clinical ratings on the functional status of patients [7].

For the past two decades, new outcome movements have placed importance on patient-based functional assessment and measurement of health-related quality of life [5, 8].

It is because the statistically significant changes of the visual analog scale (VAS) for pains, from 68 to 42, do not mean any improvement for the patient with chronic pain. The improvement of knee flexion range of 35° does not answer the questions from patients and their families.

The end result of health services should take into account the experiences, preferences, and values of the patients [9]. The discomfort described by patients should be managed properly. We have to evaluate and treat the patient not the disease.

The so-called health-related QOL became the main target of the outcome measures related to patients' values. As well, of course, it is clear that health-related QOL assessments alone are not enough. In 2001, the World Health Organization (WHO) proposed a new "health" concept called the International Classification of Functioning, Disability and Health (ICF) [10], which describes the pluralistic conception of functional disabilities [11]. Based on the ICF concept, for example, we would assume the problems associated with patients with knee osteoarthritis were as follows:

Health condition	Osteoarthritis
Impairment	Knee pain, limited joint movement, muscle weakness
Activity limitation	Difficulty in mobility-related daily activities
Participation restriction	Difficulty in participating in social life

Many items and domains in the outcome measures should be constructed to identify all the characteristics of subjects, with a view to disability and impairment. Irrgang and Anderson [12] described the process for developing such a new measure for health-related quality of life when they designed clinical research related to the knee. In many cases, several outcome measures must be combined to cover every aspect from functional impairment to quality of life.

10.8.3 BACKGROUND OF CLINICAL TRIALS

10.8.3.1 Meaning of Clinical Trial

Two related statistical factors have to be considered in order to evaluate the quality of clinical studies. This is essential for the understanding of the quality of studies based on the level of evidence. First, in clinical trials we deal with human beings as research objects. Each human being is different; response is always uncertain, backgrounds vary, and mental influence is inevitable. We should compare groups that are similarly distributed, except for the specific factor that is the target of assessment. Thus, observed differences at the assessments can be directly attributed to the study targets. Second, we mainly evaluate an extracted specimen of limited numbers and then extrapolate the conclusion toward the source population. We attribute the presumption from the sample population to the general population, using statistical inference techniques. Using the statistics observed in the investigated samples, the parameters of the source population are estimated. Depending upon these two premises, we use sampling, randomization, comparison, and other statistical methods. This is a statistical outlook of the world around us.

10.8.3.2 Estimate and Testing

When we treat a patient, we regard him or her not just as a single case but as a representative case among many patients with the same disorder. This concept of patient treatment had already been described by Hippocrates in ancient Greece. There are two important statistical concepts we should consider, probability and confidence interval.

The first topic is the probability, that is, P values that the result would arise by chance. Depending on the choice of cut-off level ($P < 0.05$ or $P < 0.01$), we apply a different level of "statistical significance" to the trial. If we find a result in the statistically significant range, we could reject the null hypothesis that there is no real difference between two groups. But a P value in the nonsignificant range just shows that either there is no difference between the groups or there are too few samples to reveal a difference, if it actually exists.

Also, if we repeated the same trial hundreds of times, we would not get the exact same result each time. We cannot omit an incidence in each trial because we usually conduct only one trial, but we could find a particular level of difference on average

with a 95% confidence interval. If we conducted a trial with a much larger sample size, we would be more likely to show whether a significant result exists or not. The answer is in the comparison of the 95% confidence limits and the clinical significant levels. Very narrow 95% confidence limits suggest definitive results and exclude any clinical obscurity in comparisons. The confidence intervals reveal whether the trials support the result and to what extent and whether any further studies need to be done to reinforce the result.

10.8.3.3 Various Biases

The results of a trial consist of three parts: (1) true value, (2) random error (imprecision) by chance, and (3) systematic error (bias) by other factors. We cannot eradicate random error, but we can eliminate the other biases by using well-designed study protocols. The goal is to suppress the biases as much as possible and improve the quality of the trial; different study designs require different steps to reduce systematic errors. In clinical epidemiology textbooks, there are various biases pointed out such as sampling bias, selection bias, and information bias.

Confounding is the distortion of an association between two factors brought about by the association of another extraneous factor; for example, the association between lung cancer and alcohol consumption is confounded by smoking. The methods of controlling confounding in the design of a study are: restrict the participants to the study, match individuals with other comparison groups, or use random allocation. In the analysis of a study, restrict once again the participants in the data to stratify the individuals into subgroups according to categories of confounding factor and use multivariate methods.

10.8.3.4 Methods of Clinical Trials

To assess the effectiveness of a certain intervention, it is necessary to compare the baseline (before state) with the result (after state) of treatment. But generally speaking, the more abnormal value often shows more improvement.

Detected improvement does not always imply that the intervention was effective. Several other factors may have contributed to the same kind of improvement; such possible factors are natural recovery, "regression to the mean" phenomenon, or the psychological effect called placebo effect. We have to compare with control to get the true change rates by test intervention.

In most clinical trials called controlled trials, there is a group receiving a certain treatment and a comparative group. An essential premise of this kind of trial is that there is genuine uncertainty as to which treatments will be best for the patient. It is this uncertainty that justifies random allocation of patients after the consent forms to enter the trial have been given. Therefore, patients should have the targeted disease and satisfy all the required conditions, that is, inclusion criteria.

10.8.3.5 Types of Clinical Studies

If there were a clinical question about a patient's management, we should perform a clinical study, no matter what the study design is. The basic structures of a study design consist of a combination of some dichotomous divisions.

1. *Experimental (Interventional) Study or Observational Study* In experimental study, a population is selected for a planned trial of a regimen whose effects are measured by comparing the outcome of the regimen in the experimental group with the outcome of another regimen in the control group. Observational study is a nonexperimental study that does not involve any intervention.
2. *Comparative or Noncomparative Design* In a comparative design, the study design has a control group to compare with the active group. In a noncomparative design, there is no control group to compare with the active group.
3. *Intergroup or Intragroup Design* This is comparison of interventional effects among the groups, or comparison in the same single group, before and after the intervention.
4. *Cross-Sectional or Longitudinal Design* This is the directionality of the study. Data are collected at a single time point or collected at two or more time points and followed up later. But some cross-sectional studies refer retrospectively to experiences in the past.
5. *Retrospective or Prospective Design* This is a part of longitudinal studies. Data are collected at baseline points and other points in the past or future. The timing of data collection is also called concurrent, historical, or mixed.

Considering these dichotomous divisions, a widely used basic study design is a parallel design, which compares the groups, or a cross-over design, which compares the before and after. In the parallel design, subjects are divided into two or more groups, each group receives each treatment concurrently, and the results of treatment among the groups are compared. In the cross-over design, subjects receive two or more treatments at stated intervals and the effect of treatment, within the same subject, is evaluated. The cross-over design has the advantage of a smaller sample size than the parallel design but requires caution to minimize the fluctuating effect of disease severity ("order effect") and maintain the stability of patients' symptoms using a "wash-out period."

The randomized controlled trial (RCT) provides the most powerful proof of treatment efficacy:

Randomized Controlled Trials Subjects are randomly allocated into two or more groups. The groups are followed up for a specified time period and assessed in terms of outcomes defined at the start of trials. As the groups, on average, are identical except for the intervention as study target, any differences in outcomes are attributed, in theory, to the interventions.

Open label Subjects know which treatment they are receiving.

More Sophisticated Randomized Controlled Trials

Single blind Subjects do not know which treatment they are receiving.

Double blind Neither subjects nor investigators know who is receiving which treatment.

10.8.3.6 Random Allocation

Blinding patients to which treatments they are receiving is thought to be essentially important because the effect of the psychological expectations of the patients are

suppressed. In drug therapy, blinding is achieved with a placebo as a control. However, in certain situations, such as a comparative trial between surgery and nonsurgery, blinding is not easily possible. Double-blind techniques for surgical treatments are tentatively proposed; the anesthesiologists and surgeons in charge will maintain silence about treatment choices and patients and the persons assessing the outcomes will not be aware of the treatment choices. Because of the technical difficulty of being unaware of allocation, we sometimes have to accept that patients would know the treatment they received.

Random allocation is done immediately after completing the original registration for a trial with informed consent. However, patients often hesitate to enter a trial because, at the time of registration, they do not know which treatment is indicated for them. Zelen proposed a new method called "randomized consent design" in which random allocation is already performed at the candidate stage of being involved in a trial. Explanation of informed consent is done only about the allocated treatment [13, 14].

10.8.4 METHODOLOGICAL ASSESSMENT FOR CLINICAL TRIALS

Because of the explosive increase in health care information, all clinicians must be careful of scientific evidence regarding efficacy of treatment methods based on available medical literature. How are we to distinguish the good from the bad among the flood of information? We need a systematic way to select medical information including RCTs.

10.8.4.1 Significance of Clinical Trials

The practice of EBM is a process of self-directed, life-long, never-ending learning for our patients. Clinicians need to have the most up-to-date clinical information. Our goal is to continuously strive to improve our medical knowledge by scientific methodology. The introduction of newer information technology and medical knowledge has improved the practice of medicine. If we can properly manage to get all of the clinical information from contemporary medical practice, we will be able to master all of the medical experiences of that same period. The collected results could be the most verified medical knowledge that we have ever had.

In the musculoskeletal field, in-patients are, in most cases, operated on, but the majority of out-patients and some in-patients are treated with a combination of conservative therapeutic methods. We need to develop a systematic strategy for the indication and selection of various treatments. The accumulated results of RCTs could be the essential medical information for the most up-to-date treatments.

10.8.4.2 Critical Evaluation of RCTs

Randomized controlled trials are now regarded as having superior clinical significance, that is, the gold standard for proper treatment, and are recognized as the primary way to provide rigorous proof of efficacy. However, considering the mixture of facts and errors found in the results of studies, we have to check each step of the trial process, even in the case of RCTs. All trials can be scored on methodological

quality according to a few criteria. Basically, these criteria consist of four main categories [15–17]:

1. Design and study population
2. Description of intervention
3. Measurement of outcomes and follow-up
4. Data presentation and analysis

In the recent analyses, these 4 categories were usually further subdivided into a set of nearly 20 items, each with a given weight [18]. Various scales for quality assessment are now available and after some modifications, those for musculoskeletal disorders have been employed in several systematic reviews and meta-analyses [19, 20]. Introduced here is one of the sophisticated examples of a criteria list for assessing quality of trials in a published article (Table 1).

10.8.5 PREPARING NECESSARY TOOLS FOR CLINICAL TRIALS

On the basis of my personal experience, the next few sections describe the minimum essential steps in conducting a clinical trial (Table 2); the reader should refer to appropriate textbooks for further reading [21–26]. Recently, this kind of information has also been available on the Internet; the Resource Center for Randomized Trials [27] is an example. The Resource Center has various activities, and one of the useful services is a Web-based library for accessing information about trials; there are checklists for trials, consent forms for participants, and patient information leaflets. Other public information on clinical trials is also provided by the U.S. government [28].

10.8.5.1 Developing Protocol

The first step in conducting a trial is developing the protocol. When the details of how the trial is to be conducted are determined, it is essential to reduce various relating biases. Farrell and Spark [29] described a protocol checklist as a detailed method to use when conducting a trial (Table 3). Developing a protocol takes a long time, sometimes up to a few years to get funding, secure ethical approval, and organize a trial collaboration team. I think the best thing for anyone planning a trial for the first time to do is to refer to some successful protocols. Many protocols are now available on the Internet. A protocol should clearly show how outcomes will be measured, data collected, and the analysis conducted.

10.8.5.2 Patients

Use of inclusion/exclusion criteria is important to make it possible for the trial results to be more reliable. Inclusion criteria are used to identify appropriate subjects, keep them safe, and able to answer the research questions. Exclusion criteria are also very important in participant recruitment to avoid unnecessary involvement of inappropriate subjects in the study.

TABLE 1 Criteria List for Assessing Methodological Quality of Trials

Heading	Subheading	Descriptor
Title		Identify the study as a randomized trial.
Abstract		Use a structured format.
Introduction		State prospectively defined hypothesis, clinical objectives, and planned subgroup or covariate analyses.
Methods	Protocol	Describe
		Planned study population with inclusion or exclusion criteria.
		Planned interventions: their nature, content, and timing.
		Primary and secondary outcome measure(s) and the minimum important difference(s), and indicate how the target sample size was estimated.
		Reasons for statistical analyses chosen, and whether these were completed on an intention-to-treat basis.
		Mechanisms for maintaining intervention quality, adherence to protocol, and assessment of fidelity.
		Prospectively defined stopping rules (if warranted).
	Assignment	Describe
		Randomization (e.g., individual, cluster, geographic).
		Allocation schedule method.
		Method of allocation concealment.
	Masking (blinding)	Describe
		Mechanism for maintaining blind and allocation schedule control.
		Evidence for successful blinding.
Results	Participant flow and follow-up	Provide a trial profile summarizing participant flow, numbers and timing of randomization assignment, interventions, and measurements for each randomized group.
	Analysis	State estimated effect of intervention on primary and secondary outcome measures, including a point estimate and measure of precision (confidence interval).
		State results in absolute numbers when feasible (e.g., 10/20, not 50%).
		Present summary data and appropriate descriptive and interferential statistics in sufficient detail to permit alternative analyses and replication.
		Describe prognostic variables by treatment group and any attempt to adjust.
		Describe protocol deviations.
	Discussion	State specific interpretations of study findings, including sources of bias and imprecision (internal validity) and discussion of external validity, including appropriate quantitative measures when possible.
		State general interpretation of the data in light of the available evidence.

Source: Modified from Machin et al. [21], with permission.

In spite of application of well-defined inclusion/exclusion criteria, it is commonly thought that about half of all clinical trials do not achieve their planned sample size [30]. As a recruitment strategy, necessary sample size is calculated to cover the expected number of dropout cases. It seems to me that the calculated number of the sample size is just for reference as an appeal to the cautious attitude of the researcher when conducting the trial.

TABLE 2 Necessary Steps to Conduct Trials in Musculoskeletal Disorders

Developing protocol	Aim and goals with using strategies
	Simple pragmatic design
	Minimal amount of data correction
	Random allocation
	Ethical considerations
Patients	Clear inclusion/exclusion criteria
	Minimal burden for patients
	Recruitment strategies
Interventions	Drug, therapeutic exercise, physical agents, orthosis, Education or care package
Outcomes	Assessment of functions and QOL, good questionnaire
	Optimal combining outcome measures
	Data management procedures
Time constrained	Limited core activity
Staff and their training	Minimal demand
	Sufficient preparation and supporting

TABLE 3 Developing a Protocol for Clinical Trial

Title
Summary
Background and rationale for the trials
Hypothesis to be tested
Primary outcome(s)
Secondary outcomes
Inclusion and exclusion criteria
Interventions to be tested
Estimated sample size
Information for patients and consent
Analyses plan, concluding dummy table
How patients will be entered into the study, concealment of allocation
Duration and methods for follow-up
Data collection, including questionnaires
Trial management
Trial supervision
Publication policy
Reference

Source: Modified from Duley and Farrell [23], with permission.

10.8.5.3 Interventions

Treatments in musculoskeletal disorders, apart from surgical interventions, are classified into "drug therapy," "therapeutic exercise," "physical agents," "orthotics and devices," "education or care packages," and "others." When these intervention methods are indicated in the experimental group, it is important how comparative methods are selected in the control group. It is impossible to use the double-blind technique with placebo in musculoskeletal disorders because the comparative control is obviously different from the experimental content for the trial participants. The patients already know which treatments have been allocated and only

the assessors are unaware of the allocation results. This "open-label method" is often indicated in this field.

10.8.5.4 Outcomes

The content of outcomes used in clinical trials should be important to patients, usually the five D's: death, disease in clinical course, discomfort in symptoms, disability in activities of daily living (ADL), and dissatisfaction in QOL. There are many de facto standards for outcome measures in musculoskeletal disorders (Table 4).

Another important point in the use of outcome measures is the timing of application. Follow-up periods for outcome measures are categorized as short term (less than 6 weeks), intermediate term (6 weeks to 1 year), or long term (more than 1 year follow-up).

In an RCT, the results basically underestimate the difference between the comparative groups because the analysis is conducted with a more conservative standpoint for assessment (an intention-to-treat principle), and that will tend to dilute the estimate of the true difference. The content of interventions often fluctuates depending upon its nature, application timing, and patients' adherence. When the patients receive the strict treatment, they can expect more obvious efficacy than that reported.

10.8.5.5 How to Choose Right Outcome Measures

As an example, there is a quick-reference book that summarizes and evaluates more than 150 outcome measures for each joint of the extremities [39]. As Suk et al.'s handbook does not include the outcomes for spinal disorders, we have to have proper outcome measures for those, including neck and back problems [40, 41] (Table 4).

The necessary information for selecting outcome measures is as follows:

1. Goal of measurement
2. Nature of measurement: questionnaire, performance rating, physical properties
3. Specific population for which the instrument was developed
4. Format of measurement: number of items, response options, minimum and maximum score
5. Issues related to feasibility: time needed to perform the measurement, required equipment, and training

For clinical researchers, outcome measurements are essential for the advancement of their studies. To assess the overall quality of an outcome measure, three major elements should be considered: content of construct, psychometric evaluation, and clinical utility.

The conventional methods of studying the dimensional structure of measures are principal component analysis and factor analysis; Cronbach's α is calculated to determine the internal consistency of the dimensions. Outcome measures usually consist of one or more domains that reflect the concept of supposed construct.

TABLE 4 Widely Used or Highly Qualified Outcome Measures for Musculoskeletal Disorders

• Low back pain	
Oswestry low back pain disability questionnaire [31]	
Roland–Morris disability questionnaire (RDQ) [32, 33]	
• Rheumatoid arthritis	
Health assessment questionnaire (HAQ) [34]	
Arthritis impact measurement scale (AIMS) [35–37]	
• Osteoarthritis (hip and knee)	
Western Ontario McMaster University (WOMAC) [38]	
(In this section of joint functions, the outcome measures that scored higher than 8 are shown)	
• Joint functions [39]	
Shoulder	Disabilities of the arm, shoulder and hand (DASH)
	Flexilevel scale of shoulder function (FLEX-SF)
	Oxford shoulder score
	Shoulder instability questionnaire
	Shoulder pain and disability index (SPADI)
	Simple shoulder test (SST)
	Upper extremity function scale
Elbow	Elbow functional assessment scale (EFA)
	Liverpool elbow score
	Upper extremity function scale
Wrist/hand	Boston questionnaire (also known as Brigham and Women's carpal tunnel questionnaire)
	Cochin rheumatoid hand disability scale
	Patients rated wrist evaluation (PRWE)
	Sequential occupational dexterity assessment (SODA)
	Upper extremity function scale
Pelvis	—
Hip	AAOS hip and knee score
	Functional recovery score
	Harris hip score
	Oxford hip score
	Western Ontario and McMaster Universities OA index (WOMAC)
Knee	AAOS hip and knee score
	Activity rating scale
	Fulkerson–Shea patellofemoral joint evaluation score
	Knee outcome survey activities of daily living scale
	Knee injury and osteoarthritis outcome score (KOOS)
	Kujala patellofemoral score (also know as the AKPS—anterior knee pain scale)
	Oxford 12-item knee questionnaire
	Western Ontario and McMaster Universities OA index (WOMAC)
Ankle	Foot health status questionnaire
Calcaneus	—

It is very important to review, integrate, and consolidate the information with measuring instruments for cross-cultural use [6].

10.8.5.6 Questionnaires

The clinical utility of an outcome measure is divided into two parts: patient friendliness (acceptability) and clinician friendliness (feasibility) [42].

If we want a questionnaire to be user friendly, for both patient and clinician, and to improve clinical utility, the questionnaire has to be refined repeatedly. The content of the questionnaire must be limited to the minimum amount of information needed and be designed to be filled out easily. Even in the case of multiple choice or fill-in-the-numeral forms, there are some problems involved in interpreting the data.

Recently, the content of questionnaires that has only been determined by so-called expert staff is no longer regarded as sufficient for assessing content validity. We have to include the opinions of patients during the development process of the questionnaires in order to check content validity and face validity (these are often used interchangeably).

10.8.6 DATA ANALYSIS AND INTERPRETATION

10.8.6.1 Sample Size

The sample size of a trial is calculated based on the estimate of primary outcome [43]. I would like to show an example of how to calculate a sample size from the difference between the two means.

We have to recognize a way of denying null hypothesis, which is always associated with false-positive and false-negative rates. The former (false positive) is known as the type I error, that is, significance level α. The latter (false negative) is the type II error β, and $1 - \beta$ is the power. The sample size is given here in the condition of two groups' comparison, and randomization in equal group numbers.

$$\text{Sample size } N = \frac{4(Z\alpha + Z\beta)^2}{\Delta^2}$$

where $Z\alpha$, $Z\beta$ are standardized normal deviates for given α, β, and Δ is standardized effect size.

$Z\alpha = Z_{1-\alpha/2}$; $\alpha = 0.05$ in two-sided, then $Z_{1-\alpha/2} = Z_{0.975} = 1.96$

$Z\beta = Z_{1-\beta}$; a power of $1 - \beta = 0.9$, then $Z_{1-\beta} = Z_{0.9} = 1.645$

$\Delta = \mu A - \mu B/\sigma$; $(\mu A - \mu B)$ is difference between the two means, and σ is standard deviation of the endpoint

If you set Δ less than 0.1, approximately 0.5, or more than 1, it means a small, moderate, or large standardized effect, respectively.

Sample size is noted in the majority of articles on clinical trials. Sample size depends on whether we want much more power to test the difference among the groups or we want to estimate the precision of the confidence interval. If we have an available working hypothesis, we can calculate the necessary sample size according to other similar studies. The result of the calculations only indicates the number required for the data analysis. Usually, we have to add a few more cases, expecting that there will be some dropouts.

10.8.6.2 Intention to Treat

Intention to treat (ITT) is a conceptual principle and is not defined as a specific technical procedure. One widely used procedure is to review the data in detail when the treatment and follow-up have been completed and all the patient information has been collected, that is, at the time that all the data has been frozen. Once allocation has been done, it is the ITT principle to analyze as if the original allocation was continued, even after the intervention itself was changed. The "last observation carry-forward" method is able to cover the deficit data when the trial has been in progress for some time (more than two check points).

However, even such principles cannot be maintained at analysis in the following cases:

1. Criteria unfit for inclusion, even after allocation
2. Complete loss to follow-up

10.8.6.3 Multiple Comparisons

We must cut down the working hypotheses to one main one. In a usual trial comparing two groups, a statistical test provides a figure of P value as the boundary line for the verdict of hypothesis. If the calculated P value for the primary endpoint is less than the predefined line (usually $P < 0.05$), we will deny the null hypothesis of no difference between the groups. However, if we use more than one endpoint, there are more than two comparisons arising (sometimes this situation becomes much more complex with multiple comparisons). In such a situation, a false-positive rate is no longer $\alpha \times 100\%$ and increases according to the number of multiple comparisons.

10.8.6.4 Post Hoc Analysis and Subgroup Analysis

It is the nature of things that the therapy-responder group has a better prognosis and the nonresponder group has a worse result. That is a typical post hoc analysis. Grouping allocation should be provided at the beginning and after disclosing the result, retrograde grouping is not permitted.

At times, clinicians like to know whether certain types of patients show more obvious benefits from interventions than others, knowing that a sample group of interest is a mixture of various types of patients. This is the subgroup analysis, for example, mild or severe cases, young or old patients. Some subgroups could easily reveal positive effects by chance, even if the overall trial results are negative. A common mistake is to mix the P value of more than 0.05 in all data-combined analysis and less than 0.05 in one or more subgroup(s). This may lead to the false conclusion that a certain subgroup has a favored result by chance, in spite of the fact that the true result is negative. When you plan a subgroup analysis at the design stage, advance adjustment in sample size should be considered.

10.8.6.5 Data Synthesis and Combined Result

If there are several numbers of articles available for review, systematic review, is regarded as an "infrastructure" of the information system supporting EBM practice.