N-of-1 trials in general practice

'tailoring treatment to individual patients'

The study presented in this thesis was performed at the Institute for Research in Extramural Medicine (EMGO Institute), VU University Medical Center, and the Department of Medical Pharmacology / Pharmacotherapy, VU University Medical Center, The Netherlands. The EMGO Institute participates in the Netherlands School of Primary Care Research (CaRe), which was acknowledged in 1995 by the Royal Netherlands Academy of Arts and Sciences (KNAW).

The study was financially supported by Leo Pharma, The Netherlands.

ISBN: 90-6464-896-4

Printing: Ponsen & Looijen b.v., Wageningen

VRIJE UNIVERSITEIT

N-of-1 trials in general practice

'tailoring treatment to individual patients'

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. T. Sminia,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Geneeskunde
op vrijdag 22 april 2005 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Anna Carolina Maria Wegman

geboren te Amsterdam

promotoren:	prof.dr. Th.P.G.M. de Vries
	prof.dr. W.A.B. Stalman
copromotor:	dr. D.A.W.M. van der Windt

Aan pa en ma

en in mooie herinnering aan Jos Wegman

Contents

Voorwoord

Een kleermaker (zoals mijn vader) maakt kleding op maat. Hij maakt van een centimeter gebruik om te bepalen wat de beste pasvorm is voor zijn klant. Een arts wil ook graag zijn/haar patiënt een behandeling geven die het "beste" voor hem/haar is. Met dit proefschrift wil ik u laten zien dat zoals een kleermaker gebruik kan maken van convectie kleding en deze aan kan passen aan de eigenschappen (bijvoorbeeld lengte en taille) en de wensen van de klant, een arts gebruik kan maken van de kennis over behandelingen die onderzocht zijn in grote studies, en met behulp van N=1 onderzoek de behandeling kan aanpassen (soort geneesmiddel, dosering, toedieningsvorm) aan de kenmerken en wensen van zijn/haar patiënt. Ik hoop dat ik uw interesse gewekt heb en wens u veel leesplezier toe.

Chapter 1

General Introduction

Background

In general practice, deciding on the optimal treatment for individual patients is one of the main concerns of the general practitioner (GP). Scientific evidence, recommendations from guidelines, medical expertise, patient preferences, and personal experiences all contribute to this decision process.[1,2] It is generally accepted that randomised controlled trials (RCTs) provide the strongest evidence for the efficacy of treatment. However, evidence from RCTs is not always available, simply because no RCTs have been carried out to investigate the specific medical problem or treatment options at issue.[3] Moreover, RCTs are designed to estimate an average treatment effect in a selected population.[4] In daily practice the GP has to determine the extent to which this average effect will apply to an individual patient in an almost unselected population. For instance, a patient who consults a GP may not be of a similar age, may have co-morbidity, may be taken additional medication, or may be interested in a different outcome, compared to the subjects studied in the related RCT.

Normally, when a GP has any doubts concerning the applicability of treatment recommendations (derived from RCTs) to a specific patient, the trial and error method is applied. This means that a particular drug will be prescribed, and will be subsequently continued if considered effective, or changed if considered not beneficial (efficacy, side effects). However, the fact that both the patient and the GP are aware of the change in treatment can influence the evaluation of its effectiveness. Individual preferences and experiences, and the expectations of both the patient and the GP can play an important role in the assessment.[3] Furthermore, the trial and error method is often biased to a false positive effect (type I error), because most disorders are self-limiting, placebo effects may occur, and the patient wants to please the GP. This may result in an over-consumption of drugs.[5] If the GP or the patient doubts the effectiveness of a certain drug treatment, and prefers an objective evaluation, the N-of-1 randomised controlled trial (N-of-1 trial) may, in certain situations be a valuable addition to RCTs and a useful tool in general practice.[6]

N-of-1 trials

N-of-1 trials belong to the family of single case experimental designs. These designs originate from psychological research,[7] and have in common that each individual is his or her own control. The individual is subsequently exposed to the active intervention(s) and the control intervention and the dependent variable is repeatedly measured.[8,9] These designs should not be confused with case studies, which describe interesting phenomena in individuals that can possibly generate hypotheses.[9]

Guyatt and his colleagues recognised the problem that physicians often encounter when they try to practice evidence based medicine: "They cannot trust their own 'uncontrolled' therapeutic trials, but neither can they often look to large-scale randomised trials for definitive treatment recommendations".[3] To find a possible solution for this dilemma, several single case experimental designs were tested, and the N-of-1 trial was found to be the most efficient and feasible way to evaluate drug treatment in medical care.[3]

The N-of-1 trial is a multiple crossover trial in one patient, consisting of a series of pairs of treatment periods. The patient is his or her own control, and receives during one period of each pair the active treatment, and during the other period the control intervention or the placebo. The sequence of treatments is randomised within each pair of treatment periods (Figure 1). The patient, the physician, and the researcher are all blinded for the sequence of treatments.[3]

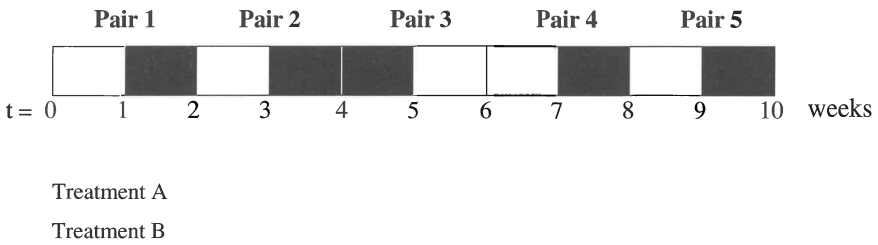

Figure 1. Example of randomisation schedule for one patient receiving 5 pairs of treatments, each consisting of one week of treatment A and one week of treatment B

However, N-of-1 trials cannot be conducted for all medical problems. From a methodological point of view, a number of conditions have to be fulfilled with regard to the medical problem and the interventions. First, the medical problem has to be chronic; secondly, the treatment effects of the interventions should have a rapid onset and stop acting soon after discontinuation to prevent carry-over effects (i.e. the effect of one treatment should have disappeared when the patient is given the other treatment); and thirdly, it is sensible to conduct N-of-1 trials only if there are real doubts about the treatment policy.[6]

Research topics in this thesis

For GPs, the N-of-1 method is a relatively new research design, and reports of N-of-1 trials in general practice are relatively scarce. Consequently, the value of N-of-1 trials for patient care in general practice is not yet clear. To assess the medical problems in general practice for which N-of-1 trials could be most suitable and relevant, we carried out a small survey among seven GPs. Of the most prevalent medical problems in general practice in The Netherlands, they considered osteoarthritis (OA) and sleep disturbances to be the most suitable and relevant for conducting series of N-of-1 trials (Appendix).

OA is a common musculoskeletal disorder. The prevalence of OA increases with age to over 50% in the knee and approximately 20% in the hip in women aged 80 years and older. In elderly men, the prevalence of OA of the knee is approximately 25%, and the hip is affected in more than 10%.[10] Pain relief and improvement of functional disability is the primary goal of the treatment, which often needs to be continued for long periods of time.[11,12] In general practice, non-steroidal anti-inflammatory drugs (NSAIDs) are prescribed in 35% to 78% of patients with OA.[13-16] However, it is generally known that NSAIDs are associated with gastrointestinal complications, whereas paracetamol is associated with very few side effects.[17-19] Therefore, the majority of early clinical guidelines for OA recommend paracetamol as first-line drug treatment.[11,12,20-22] These recommendations were strengthened by the results of two RCTs demonstrating no significant benefit of NSAIDs over paracetamol.[23,24] Recently, additional evidence has appeared, and existing guidelines have been updated [25] and new ones have been

developed [26-28]. However, the validity and interpretation of the available evidence, and the quality and content of guidelines for the pharmacological management of OA has been debated.[29-35] Furthermore, the relative effectiveness of NSAIDs, compared to paracetamol, seems to vary across patients, with approximately 57% preferring NSAIDs, 20% preferring paracetamol, and 22% reporting no preference.[36] N-of-1 trials may provide an answer for individual patients as to whether it is better to take NSAIDs or paracetamol.

Benzodiazepines are frequently prescribed for sleep disturbances.[37] In 1998, the annual prevalence of benzodiazepine use was estimated at 12.2% in The Netherlands. At least one third of all users were long-term users.[38] However, there is considerable doubt about whether benzodiazepines are still effective when taken for a prolonged period of time.[39-41] Additionally, benzodiazepine use is associated with adverse cognitive effects and an increased risk of motor vehicle accidents, falls and fractures, especially among the elderly.[42-44] Furthermore, benzodiazepines may interact with other medication, such as antidepressants and antiepileptic drugs, resulting in an intensified hypnotic effect.[45] Therefore, in practice guidelines for the management of insomnia it is recommended to prevent or stop long-term use of benzodiazepines.[39,41] N-of-1 trials may help to determine whether the use of hypnotics can be successfully reduced in individual patients without the loss of quality and quantity of sleep.

Objectives of this thesis

The overall aims of the research presented in this thesis were (1) to study whether N-of-1 trials can be complementary to evidence derived from RCTs for decisions regarding drug treatment for individual patients, (2) to study the practical and methodological difficulties encountered in N-of-1 trials, and (3) to study the applicability and feasibility of N-of-1 trials in the treatment of individual patients in general practice. For the sake of clarity, in this thesis an N-of-1 trial is considered as a tool for the GP to determine a treatment policy for an individual patient in an evidence based way, whenever the GP or the patient has doubts about the treatment policy. However, to study the usefulness, the difficulties, the applicability and the feasibility of this 'tool' for the GP, we conducted two *series* of

N-of-1 trials. These *series* of N-of-1 trials were initiated by our research group and not, in the first instance, by a GP or a patient.

The overall aims of the thesis led to the following study objectives:

- to systematically summarise the available evidence on the efficacy of non-steroidal anti-inflammatory drugs (NSAIDs) compared to paracetamol, and to compare the quality and content of clinical guidelines regarding the pharmacological treatment of osteoarthritis of the hip or knee. (Chapter 2, Design: a systematic review of randomised controlled trials and guidelines)

- to investigate for individual patients who have been using NSAIDs regularly, whether paracetamol is as effective as NSAIDs in the treatment of pain and disability related to osteoarthritis of the hip or knee. (Chapter 3, Design: a series of N-of-1 trials in general practice)

- to investigate for individual patients whether placebo is as effective as temazepam, or, in some patients, whether 10 mg is as effective as 20 mg temazepam, and to investigate whether presenting the personal results of the N-of-1 trial to each individual patient influences their future use of temazepam. (Chapter 4, Design: a series of N-of-1 trials in general practice)

- to describe the difficulties we encountered during the series of N-of-1 trials in patients with osteoarthritis of the hip and knee and the series of N-of-1 trials in long-term users of temazepam, and to discuss our solutions to these difficulties. (Chapter 5, Design: a descriptive study)

- to study the barriers and facilitators experienced by general practitioners with regard to the applicability and feasibility of N-of-1 trials in general practice. (Chapter 6, Design: a qualitative study based on semi-structured interviews)

Finally, the main results and conclusions of this thesis, and a discussion of their implications for future research, general practice, and patients are presented in Chapter 7.

References

1. Sacket DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71-2.
2. Elwyn G, Edwards A, Wensing M, Hibbs R, Wilkinson C, Grol R. Shared decision making observed in clinical practice: visual displays of communication sequence and patterns. J Eval Clin Pract, 2001;7:211-21.
3. Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determing optimal therapy – randomized trials in individual patients. N Engl J Med 1986;314:889-92.
4. Altman DG. Practical statistics for medical research. London: Chapman & Hall, 1991:1-9.
5. Johannessen T, Controlled trials in single subjects: 1 Value in clinical medicine. BMJ 1991;173-4.
6. Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D, Keller J. A clinician' s guide for conducting randomized trials in individual patients. CMAJ 1988;139:497-503.
7. Barlow DH, Hersen M. The single-case in basis and applied research: an historical perspective. In: Barlow DH, Hersen M, eds. Single case experimental designs – strategies for studying behavior change. 2nd ed. New York: Pergamon Press,1984:1-31.
8. Barlow DH, Hersen M. General issues in a single-case approach. In: Barlow DH, Hersen M, eds. Single case experimental designs – strategies for studying behavior change. 2nd ed. New York: Pergamon Press,1984:32-31.
9. Bryson-Brockmann W, Roll D. Single-case experimental designs in medical education: an innovative research method. Acad Med 1996;71:78-85.
10. Van Saase JL, Van Romunde LK, Cats A, Vandenbroucke JP, Valkenburg HA. Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations. Ann Rheum Dis 1989;48:271-80.
11. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR et al. Guidelines for the medical management of osteoarthritis. Part I. Osteoarthritis of the hip. American College of Rheumatology. Arthritis Rheum 1995;38:1535-40.
12. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR et al. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. American College of Rheumatology. Arthritis Rheum 1995;38:1541-6.
13. Bierma-Zeinstra SM, Lipschart S, Njoo KH, Bernsen R, Verhaar J, Prins A et al. How do general practitioners manage hip problems in adults? Scand J Prim Health Care 2000;18:159-64.

14. Traynor V, Britt H, Neary S, Sayer GP, Charles J, Meza RA. The management of osteoarthritis in general practice. Results from the Australian Morbidity and Treatment Survey, 1990-1991. Aust Fam Phys 1994;23:1971-8.
15. Mamlin LA, Melfi CA, Parchman ML, Guttierez B, Allen DL, Katz BP et al. Management of osteoarthritis of the knee by primary care physicians. Arch Fam Med 1998;7:563-7.
16. Mazzucca SA, Brandt KD, Katz BP, Dittus RS, Freund DA, Lubitz R et al. Comparison of general internists, family physicians, and rheumatologists managing patients with symptoms of osteoarthritis of the knee. Arthr Care Res 1997;10:289-99.
17. Anonymous. Do certain NSAIDs give less cause for gastrointestinal side effects than others? [In Dutch: Geven bepaalde NSAID' s minder aanleiding tot gastrointestinale bijwerkingen dan andere?] GeBu 1994;28:85-6.
18. MacDonald TM, Morant SV, Robinson GC, Shield MJ, McGilchrist MM, Murray FE, McDevitt DG. Association of upper gastrointestinal toxicity of non-steroidal anti-inflammatory drugs with continued exposure: cohort study. BMJ 1997;315:1333-7.
19. Van der Kuy A et al. Analgesics, antirheumatic agents and antigout agents. [In Dutch: Analgetica, antirheumatica en jichtmiddelen.] In: Van der Kuy A, ed. Farmacotherapeutisch Kompas 1999. Amstelveen: ZiekenfondsRaad, 1999:914-69.
20. Scott DL. Guidelines for the diagnosis, investigation and management of osteoarthritis of the hip and knee. Report of a Joint Working Group of the British Society for Rheumatology and the Research Unit of the Royal College of Physicians. J R Coll Physicians Lond 1993;27:391-6.
21. Bijl D, Dirven-Meijer PC, Opstelten W, Raaijmakers AJ, Scholten RJPM, Eizinga WH et al. General practice guidelines for non-traumatic knee complaints in adults. [In Dutch: NHG-standaard Niet-traumatische Knieproblemen bij volwassenen.] Huisarts Wet 1998;41:344-50.
22. Eccles M, Freemantle N, Mason J. North of England evidence based guideline development project: summary guideline for non-steroidal anti-inflammatory drugs versus basic analgesia in treating the pain of degenerative arthritis. BMJ 1998;317:526-30.
23. Williams HJ, Ward JR, Egger MJ, Neuner R, Brooks RH, Clegg DO et al. Comparison of naproxen and acetaminophen in a two-year study of treatment of osteoarthritis of the knee. Arthritis Rheum 1993;36:1196-206.
24. Bradley JD, Brandt KD, Katz BP, Kalasinki LA, Ryan SI. Comparison of an antiinflammatory dose of ibuprofen, an analgesic dose of ibuprofen, and acetaminophen in the treatment of patients with osteoarthritis of the knee. N Engl J Med 1991;325:87-91.

25. American College of Rheumatology Subcommittee on osteoarthritis guidelines. Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. Arthritis Rheum 2000;43:1905-15.
26. Holbrook AM (Chair) for Ontario Musculoskeletal Therapy Review Panel. Medical treatment guidelines for the treatment of osteoarthritis, rheumatoid arthritis, and acute musculoskeletal injury. Toronto: Ontario Ministry of Health and Long-term Care, 2000.
27. Puhl W, Bernau A, Böhle E. Brune K, Gerhardt P, Greitemann B et al. Ambulatory diagnosis and therapy of gonarthrosis. [in German: Ambulante diagnostik und Therapie der Gonarthrose.] Z Orthop 2000;138:85-93.
28. Pendleton A, Arden N, Dougados M, Doherty M, Bannwarth B, Bijlsma JW et al. EULAR recommendations for the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis 2000;59:936-44.
29. Wolfe F, Zhao S, Lane N. Preference for nonsteroidal antiinflammatory drugs over acetaminophen by rheumatic disease patients. Arthritis Rheum 2000;43:378-85.
30. Moskowitz RW. The role of anti-inflammatory drugs in the treatment of osteoarthritis: a United States viewpoint. Clin Exp Rheumatol 2001;19 Suppl 25:S3-8.
31. Hungin APS, Kean WF. Non-steroidal anti-inflammatory drugs: overused or underused in osteoarthritis? Am J Med 2001;110:8S-11S.
32. Jawad AS. EULAR recommendations for the management of knee osteoarthritis. Ann Rheum Dis 2001;60:540.
33. Felson DT. The verdict favors nonsteroidal antiinflammatory drugs for treatment of osteoarthritis and a plea for more evidence on other treatments. Arthritis Rheum 2001;44:1477-80.
34. Dieppe PA. Concerns about the methodology used in developing the 2000 update of the American College of Rheumatology recommendations for management of hip and knee osteoarthritis.[letter] Arthritis Rheum 2001;44:2450-1.
35. Brandt KD. A critique of the 2000 update of the American College of Rheumatology recommendations for management of osteoarthritis.[letter] Arthritis Rheum 2001;44:2451-5.
36. Pincus T, Koch GG, Sokka T, Lefkowith J, Wolfe F, Jordan JM et al. A randomized, double-blind, crossover clinical trial of diclofenac plus misoprostol versus acetaminophen in patients with osteoarthritis of the hip or knee. Arthritis Rheum 2001;44:1587-98.
37. Kruse WH. Problems and pitfalls in the use of benzodiazepines in the elderly. Drug Saf 1990;5:328-44.

38. Gorgels WJMJ, Oude Voshaar RC, Mol AJJ, Breteler MHM, Lisdonk EH van de, Zitman FG. Long-term use of benzodiazepines. [In Dutch: Het langdurig gebruik van benzodiazepinen.] Ned Tijdschr Geneeskd 2001;145:1342-6.
39. Knuistingh Neven A, Graaff WJ de, Lucassen PLBJ, et al. National practice guideline for insomnia and hypnotics (Dutch College of General Practitioners). [In Dutch: NHG-Standaard Slapeloosheid en Slaapmiddelen.] In: Rutten GEHM, Thomas S (eds.). NHG-Standaarden voor de huisarts. Utrecht: Bunge 1993:264-77.
40. Lader MH. Limitations on the use of benzodiazepines in anxiety and insomnia: are they justified? European Neuropsychopharmacology 1999;9(Suppl 6):S399-S405.
41. Blow FC (consensus panel chair). Substance Abuse Among Older Adults. Treatment Improvement Protocol (TIP) Series 26. Rockville: U.S. Department of Health and Human Services (DHHS), 1998.
42. Thomas RE. Benzodiazepine use and motor vehicle accidents. Systematic review of reported association. Can Fam Physician 1998;44:799-808.
43. Ray WA, Fought RL, Decker MD. Psychoactive Drugs and the Risk of Injurious Motor Vehicle Crashes in Elderly Drivers. Am J Epidemiol 1992;136:873-83.
44. Herings RM, Stricker BH, Boer A de, Bakker A, Stuurmans F. Benzodiazepine and the risk of falling leading to femur fractures. Dosage more important than elimination half-life. Arch Intern Med 1995;155:1801-7.
45. Van Loenen AC, ed. Central nervous system (mental disorders). [In Dutch: Centrale zenuwstelsel (psychische aandoeningen)] In: Van Loenen AC, ed. Farmacotherapeutisch Kompas 2002. Amstelveen: College voor zorgverzekeringen (CVZ), 2001:53-135.

Chapter 2

Non-steroidal anti-inflammatory drugs or paracetamol for osteoarthritis of the hip or knee? A systematic review of evidence and guidelines

Published as:
Wegman ACM, Van der Windt DAWM, Van Tulder MW, Stalman WAB, De Vries ThPGM. Nonsteroidal antiinflammatory drugs or acetaminophen for osteoarthritis of the hip or knee? A systematic review of evidence and guidelines. J Rheumatol 2004;31:344-54.

Abstract

Objectives: The interpretation of available evidence on the relative efficacy of non-steroidal anti-inflammatory drugs (NSAIDs) and paracetamol in osteoarthritis (OA) has recently been debated. This systematic review summarises the available evidence on the efficacy of NSAIDs compared to paracetamol, and compares the quality and content of clinical guidelines regarding the pharmacological treatment of OA.

Methods: Published reports of randomised controlled trials (RCTs) and clinical guidelines were identified by a systematic search of bibliographic databases and relevant websites. The quality of RCTs was assessed by 2 reviewers independently using a standardised checklist. Data from these RCTs were used to calculate pooled differences between groups for pain and disability. The methodology of identified guidelines was appraised using the AGREE (Appraisal of Guidelines for Research and Evaluation) instrument.

Results: The search strategy resulted in the identification of 5 RCTs. Statistical pooling of data from 3 trials with adequate methods and sufficient data presentation resulted in a pooled standardised mean difference for general pain of 0.33 (95% CI 0.15 to 0.51), indicating a small effect in favour of NSAIDs. Pooled estimates for other outcome measures were smaller. Three of the 9 identified guidelines satisfied more AGREE criteria than others, particularly regarding rigor of development. Stakeholder involvement, applicability, and editorial independence were poorly described in most guidelines. The content of recommendations regarding the use of NSAIDs or paracetamol was fairly consistent.

Conclusions: Paracetamol is often effective in OA and is associated with fewer adverse reactions than NSAIDs. Available evidence supports the recommendations of recent guidelines to use paracetamol as initial therapy for OA in addition to non-pharmacological interventions. Further research is needed to establish the efficacy of NSAIDs or paracetamol in relevant subgroups of patients. We agree with guidelines that it is important that treatment is tailored to individual patients taking into account the severity of symptoms, previous use of paracetamol, and the patient's knowledge, expectations, and preferences.

Introduction

Osteoarthritis (OA) is a major cause of pain, disability, and health care use in the middle-aged and elderly population. Estimates of its prevalence depend on variations in definition, but OA is thought to affect more than 10 to 12% of the population.[1,2] With the increasing number of elderly, the prevalence and impact of OA is expected to increase over the next decades. Pain relief and improvement of functional disability is the primary goal of treatment, which often needs to be continued for long periods of time.[3,4]

Most patients with symptoms of OA are treated by general practitioners. Non-pharmacological interventions, such as patient education, exercise, or occupational therapy are the mainstay of treatment, but oral medication is often an important element of therapy.[3-5] In general practice, non-steroidal anti-inflammatory drugs (NSAIDs) are prescribed in 35 to 78% of patients with OA.[6-9] Given the relatively high risk of gastrointestinal (GI) complications, NSAIDs should not be prescribed for long periods of time, and only reluctantly in patients with increased risk of serious side effects.[10-12] A simple analgesic, such as paracetamol (= acetaminophen), has been reported to be well tolerated with few common side effects. Use of paracetamol has been associated with chronic renal failure. However, causal relations have not been established between paracetamol and chronic renal failure. Fored, *et al* discuss the fact that symptoms of diseases that predispose patients to renal failure may lead to an increased use of analgesics, thus possibly introducing a protopathic (reverse causality) bias.[13] Hepatic toxicity occurs rarely in doses below 4 g/day, but patients with excessive alcohol consumption may be at increased risk for this adverse event.[14,15] Finally, paracetamol does not affect platelet function, but can interact with warfarin, and may influence anticoagulation.[16]

Given the relatively low risk of side effects, the majority of early clinical guidelines for OA recommend paracetamol as first-line therapy.[3,4,17-19] These recommendations were strengthened by the results of 2 randomised clinical trials (RCTs) demonstrating no significant benefit of NSAIDs over paracetamol.[20,21] Recently, additional evidence has appeared, and guidelines have been updated [22] or newly developed [5,23,24]. However, the validity and interpretation of available evidence and the quality and content of guidelines

for the pharmacological management of OA have been debated.[25-31] Criticisms include the completeness of the literature search, interpretation of available evidence, differentiation between opinion and evidence, and the presence of unbalanced or biased recommendations.

Given the often prolonged pharmacotherapy, with associated side effects and costs, it is important to obtain good insight into the available evidence and recommendations regarding pharmacotherapy. We systematically evaluated the available evidence from RCTs on the short-term and long-term efficacy of NSAIDs compared to paracetamol for OA of the hip or knee. We also critically appraised the quality of guidelines on the management of OA, and compared the content of recommendations in these guidelines regarding treatment of OA with either NSAIDs or paracetamol.

Materials and methods

Search strategies

Publications on RCTs were retrieved by a computerised search of Medline, Embase, and the Cochrane Database (until December 2001). For the identification of RCTs the first 2 stages of the search strategy of the UK Cochrane Centre were used.[32] This strategy was combined with a search for OA, hip, knee, non-steroidal anti-inflammatory drugs, paracetamol, and analgesics. For the identification of guidelines a systematic search in Medline (until December 2001) was conducted using the search terms osteoarthritis, hip, knee, and guidelines. Further, several sites on the Internet were screened for publications on guidelines. Finally, the references of all retrieved articles, including systematic reviews, were screened for potentially relevant publications.

Selection of available evidence (RCTs) and guidelines

Citations from computerised databases were blinded for authors, affiliation, and source. For the review of available evidence on the efficacy of NSAIDs compared to paracetamol, we included publications that met the following conditions: (1) comparison of NSAIDs with paracetamol; (2) patients with pain and/or disability related to OA of the hip or knee; (3) at least one of the following outcome measures: overall change, pain, or

disability; (4) random allocation of interventions; and (5) publication as a full report (letters and abstracts were excluded). Language restrictions were not used.

For the review of guidelines on the pharmacological management of OA, the following selection criteria were used: (1) development of the guideline by a working group of experts (representatives of a professional group, not individual authors); and (2) recommendations are given on the pharmacological management of hip or knee OA. Systematic reviews or narrative reviews were not included. Language restrictions were not used.

Quality assessment of available evidence

Two reviewers (DvdW and MvT) independently scored the methodological quality of each trial using the internal validity criteria of the Amsterdam-Maastricht Consensus List for Quality Assessment.[33] In this checklist much emphasis is put on an adequate randomisation procedure and sufficient blinding (5 out of 11 criteria). Other criteria in the checklist refer to prognostic similarity of intervention groups at baseline, control for co-interventions, compliance, length of follow-up, dropout rate, and intention-to-treat analysis. Disagreements between the reviewers were identified and discussed during a consensus meeting. A total score for methodological quality was calculated by summing the total number of positively scored criteria (maximum score 11 points).

Data extraction and analysis of available evidence

Details on characteristics of study population, interventions, outcome measures, follow-up, side effects, and results were extracted for each randomised trial. For outcomes on a dichotomous scale the differences in proportions between study groups were computed (risk difference), together with the 95% confidence intervals (CI). Subsequently, the number needed to treat (NNT) was calculated.[34] For outcomes evaluated on a continuous or interval scale, standardised mean differences (SMD) were computed as the difference between the mean change in outcome since baseline in the compared groups, divided by the pooled standard deviation of change scores. A SMD of 0.2 can be considered to be a

small effect, 0.5 moderate, and > 0.8 a large effect.[35] A positive NNT or SMD indicated superior effects of NSAIDs, a negative NNT or SMD superior effects of paracetamol.

Statistical pooling of results was considered if there was sufficient clinical homogeneity regarding study populations, interventions, and outcome measures. A chi-square test was used to detect statistical heterogeneity of trial results. In case of statistical heterogeneity ($p < 0.10$), potential sources of heterogeneity were explored, including differences among trials in type and dosage of NSAIDs, dosage of paracetamol, severity of OA, or aspects of validity (dropout rate, intention-to-treat analysis, and use of escape medication). Pooled estimates of outcome (random effects model) were computed for homogeneous subgroups of trials.[36,37]

Quality assessment of guidelines

The AGREE (Appraisal of Guidelines for Research and Evaluation) instrument was used to critically assess the design of guidelines on the pharmacological management of OA.[38] This checklist includes items on the scope and purpose of the guideline, stakeholder involvement, rigor of development, clarity and presentation, applicability, and editorial independence (Table 1). Each item was scored by 2 reviewers independently (AW and DvdW), and then consensus was obtained on the final score. The scoring system of the AGREE instrument (4-point ordinal scale) was collapsed into a 2-point scale (+ = agree; - = disagree), and a category was added that could be used to indicate absence of sufficient information on a specific item (? = unclear).

The following details were extracted and compared across guidelines: year of publication, country, target population, number of RCTs available at the time of guideline development, evidence used for recommendations, methods used to formulate recommendations, and content of recommendations.

Table 1. Criteria for appraising the quality of clinical guidelines (AGREE criteria) [38]

Each item was scored as + (agree), - (disagree), or ? (insufficient information).

Scope and purpose	
1	The overall objective of the guideline is specifically described.
2	The clinical question covered by the guideline is specifically described.
3	The patients to whom the guideline is meant are specifically described.
Stakeholder involvement	
4	The guideline development group includes individuals from all relevant professional groups.
5	The patients' view and preferences have been sought.
6	The target users of the guideline are clearly defined.
7	The guideline has been piloted among end users.
Rigour of development	
8	Systematic methods were used to search for evidence.
9	The criteria for selecting the evidence are clearly described.
10	The methods used for formulating the recommendations are clearly described.
11	The health benefits, side effects, and risks have been considered in formulating the recommendations.
12	There is an explicit link between the recommendations and the supporting evidence.
13	The guideline has been externally reviewed by experts prior to its publication.
14	A procedure for updating the guideline is provided.
Clarity and presentation	
15	The recommendations are specific and unambiguous.
16	The different options for management of the condition are clearly presented.
17	Key recommendations are easily identifiable.
18	The guideline is supported with tools for application.
Applicability	
19	The potential organisational barriers in applying the recommendations have been discussed.
20	The potential cost implications of applying the recommendations have been considered.
21	The guideline presents key review criteria for monitoring and/or audit purposes.
Editorial independence	
22	The guideline is editorially independent from the funding body.
23	Conflicts of interest of guideline development members have been recorded.

Results

Search results

The search for RCTs yielded over 1500 citations from Medline, Embase, the Cochrane Controlled Trials Register, and reference checking. Nearly all publications were excluded after assessment of titles and abstracts, as most concerned non-randomised studies or trials comparing 2 different types of NSAIDs. For 14 citations the full publications were retrieved. Seven of these 14 publications were excluded as they compared paracetamol with placebo [39] or another analgesic [40-42], used paracetamol in combination with another drug [43], used a mixed patient population without separate presentation of results for OA [44], or was only available as an abstract.[45] Seven publications, describing the results of 5 different RCTs, were finally included in the review and subjected to quality assessment and data extraction.[20,21,46-50]

The search for guidelines resulted in 211 citations from Medline and 6 from the Internet. Nine guidelines (10 publications) were included in the review.[3-5,17-19,22-24,51] Excluded publications often concerned narrative or systematic reviews or did not specifically concern the pharmacological management of OA.

Available evidence on the efficacy of NSAIDs compared to paracetamol

Table 2 presents details of the selected trials with respect to quality score, study population, interventions, and results. The quality score varied between 5 and 8 points out of a maximum score of 11. Methodological shortcomings often concerned inadequate control for co-interventions or insufficient length of follow-up (< 3 months). A longer follow-up period was only carried out by Williams, *et al* [20]. Although procedures for randomisation and blinding were considered adequate in the trials by Solomon and Abrams [49] and Wojtulewski, *et al* [50], the publications refer to small crossover trials (42 and 24 participants, respectively) with inadequate analysis and presentation of results.

Table 2. Details of RCTs comparing NSAIDs and paracetamol (PCM): quality score, study population, interventions, and results (including SMD and 95% CI)

Study	Quality Score	Study population	Interventions (n)	Follow-up	Overall change (95% CI)	Pain on Motion (95% CI)	Pain at Rest/ General Pain (95% CI)	Disability (95% CI)
Pincus [46]	8	OA hip or knee (n = 227, 71% women). Mean age 62 years, duration of symptoms unclear. Washout 3 to 7 days before both treatment periods, increase of pain during washout.	i: diclofenac 2 x 75 mg + 200 µg misoprostol daily (112) ii: PCM 8 x 500 mg daily (115) (6 weeks, crossover)	at 6 weeks (12 weeks, after cross-over)	Investigator estimate (0-100), mean change (SD): i: -9.3 (18.5) ii: -3.6 (19.0) SMD = 0.30 (0.04, 0.57)		HAQ-VAS (0-100): mean change (SD) i: -20.8 (25.7) ii: -13.1 (28.6) SMD = 0.28 (0.01, 0.55)	WOMAC joint (0-100): mean change (SD) i: -12.2 (21.6) ii: -6.6 (18.0) SMD = 0.28 (0.01, 0.55) HAQ-ADL (0-3) mean change (SD) i: -0.16 (0.31) ii: -0.08 (0.32) SMD = 0.25 (-0.01, 0.52)
Williams [20]	7	OA knee (n = 178, 75% women). Median duration of symptoms 39 (i) and 71 (ii) months, mean age 59 years. No NSAIDs in preceding 3 months. No washout preceding trial.	i: naproxen: 2 x 375 mg daily (90) ii: PCM 4 x 650 mg daily (88) duration: 2 years	every 6 weeks until withdrawal (2 years)	Physician assessment (1-5): mean change (SD) at 6 weeks: i: 0.338 (0.647) ii: 0.274 (0.672) SMD = 0.11 (-0.22, 0.43)	Pain on motion (0-10): mean change (SD) at 6 weeks i: 1.027 (2.481) ii: 0.703 (2.062) SMD: 0.14 (-0.18,0.47)	Pain at rest (0-10): mean change (SD) at 6 weeks i: 0.909 (2.2721) ii: 0.100 (2.386) SMD = 0.35 (0.02,0.67)	50 ft walk time (s) mean change (SD): at 6 weeks i: 1.081 (3.884) ii: 0.443 (3.940) SMD = 0.16 (-0.16,0.49)
					at 2 years (n = 62): i: 0.2 (0.60), ii: 0.3 (0.90) SMD = -0.13 (-0.64, 0.37)	at 2 years (n = 62): i: 2 (3.2), ii: 1 (2.9) SMD = 0.32 (-0.18, 0.83)	at 2 years (n = 62): i: 2 (2.7), ii: 1 (2.6) SMD = 0.37 (-0.13, 0.88)	at 2 years (n = 62): i: 3 (3.2), ii: 0 (3.3) SMD = 0.91 (0.38, 1.44)

CI: confidence interval, NNT: number needed to treat, SMD: standardised mean difference, HAQ: health assessment questionnaire, VAS: visual analogue scale, PCM: paracetamol.

Table 2. Continued

Study	Quality score	Study Population	Interventions (n)	Follow-up	Overall Change (95% CI)	Pain on Motion (95% CI)	Pain at Rest/ General Pain (95% CI)	Disability (95% CI)
Bradley [21,47,48]	6	OA knee (n = 184 patients, 74% women). Mean duration of symptoms 9 years, mean age 56 years. Washout 3 to 7 days, all had pain after washout	i: ibuprofen 4 x 600 mg daily (64) ii: ibuprofen 4 x 300 mg daily (65) iii: PCM 4 x 1000 mg daily (66) duration: 4 weeks	at 4 weeks	Improved according to physician (%, 95% CI): i: 38% (26 to 50) ii: 44% (32 to 57) iii: 37% (24 to 49) i vs iii: 1% (-16 to 18%); NT=100 ii vs iii: 7% (-9 to 25%); NNT=14	Walking pain (0-3) mean change (SD): i: 0.45 (0.96) ii: 0.31 (0.81) iii: 0.13 (0.75) SMD i vs iii: 0.37 (0.01,0.73) SMD ii vs iii: 0.23 (-0.13, 0.59)	Pain at rest (0-3) mean change (SD): i: 0.40 (1.04) ii: 0.33 (0.68) iii: 0.06 (0.71) SMD i vs iii: 0.38 (0.02, 0.74) SMD ii vs iii: 0.39 (0.03, 0.75)	50 ft walk time (sec) mean change (SD): i: 0.7 (3.4) ii: 0.5 (3.8) iii: 0.5 (2.0) SMD i vs iii: 0.07 (-0.29, 0.43) SMD ii vs iii: 0.00 (-0.36, 0.36)
Solomon & Abrams [49]	6	OA knee (n = 42). Patient characteristics (age, sex, duration OA) unknown. Washout 7 days before trial.	i: ketoprofen 200 mg daily ii: PCM 6000 mg daily duration: 7 days (cross-over)	at 2 weeks (after cross-over)	Patient preference: ketoprofen: 25 (59%) PCM: 12 (29%) no preference: 5 (12%)	Not presented		Not presented
Wojtulewski [50]	5	OA hip or knee (n = 24). Patient characteristics are not described. Washout not described.	i: fenoprofen 600 mg ii: PCM 990 mg iii: placebo single dose (cross-over)	at 6 hours	Not presented	Pain relief (0-3) mean: i: 1.4, ii: 1.2, iii: 1.0		Not presented

CI: confidence interval, NNT: number needed to treat, SMD: standardised mean difference, HAQ: health assessment questionnaire, VAS: visual analogue scale, PCM: paracetamol.

The 3 other trials included larger patient populations (ranging between 178 and 227 participants), and provided detailed information on measures of pain, disability, and overall change.

Pincus, *et al* [46] reported statistically significant differences in favour of diclofenac (150 mg plus misoprostol) compared to paracetamol (4 g) for overall change, general pain (Health Assessment Questionnaire), and disability [Western Ontario and McMaster University (WOMAC) OA Index target joint score]. These differences were further reinforced in the crossover period. SMD ranged between 0.28 and 0.30. The trial by Bradley, *et al* [21] showed statistically significant benefits only for measures of pain, with SMD ranging between 0.37 and 0.39. The trial by Williams, *et al* [20] demonstrated statistically significant differences in favour of naproxen for pain at rest only (SMD = 0.35). The results of long-term follow-up are difficult to interpret due to considerable dropout after 2 years.

Additional subgroup analyses by Pincus, *et al* [46] revealed that greater improvements with NSAIDs (compared to paracetamol) were only found for patients with at least moderate pain or disability. These results were not confirmed by secondary, post hoc analyses of the trial by Bradley, *et al* [21], in which no association between level of pain and response to NSAIDs was found.[48]

Table 3 shows the results of pooled analyses of the 3 trials that were of adequate methodological quality, and provided sufficient data to enable meta-analysis.[20,21,46] All analyses showed sufficient clinical homogeneity to allow statistical pooling. Pooled estimates of differences between NSAIDs and paracetamol were statistically significant, yet small, in favour of NSAIDs (high or low dose) for general pain or pain at rest (SMD = 0.33, 95% CI 0.15-0.51). These results are presented in more detail in Figure 1. For the other outcome measures the differences were smaller. The size of pooled estimates for differences in pain on motion, functional disability, and overall change (physician assessment) ranged between 0.18 and 0.24 (Table 3).

Table 3. Pooled differences between NSAIDs and paracetamol for osteoarthritis of the hip or knee (4 to 6 weeks follow-up)

Outcome measure	Included trials [references]	Test for statistical homogeneity (χ^2)	Pooled SMD (95% CI) (random effects model)
Overall change (physician assessment)	2 [20,46] †	0.85, p = 0.36	0.22 (0.02, 0.43)
Pain on motion*			
- comparison with high dose ibuprofen	2 [20,21]	0.83, p = 0.36	0.24 (0.00, 0.48)
- comparison with low dose ibuprofen		0.12, p = 0.73	0.18 (-0.06, 0.42)
Pain at rest / general pain*			
- comparison with high dose ibuprofen	3 [20,21,46]†	0.20, p = 0.90	0.33 (0.15, 0.50)
- comparison with low dose ibuprofen		0.23, p = 0.89	0.33 (0.15, 0.51)
Functional disability *			
- comparison with high dose ibuprofen	3 [20,21,46] †	0.90, p = 0.64	0.19 (0.01, 0.37)
- comparison with low dose ibuprofen		1.54, p = 0.46	0.18 (0.00, 0.35)

* The trial by Bradley et al. [21] includes three intervention groups; paracetamol is compared to high dose ibuprofen (2400 mg daily) and low dose ibuprofen (1200 mg daily). † For the trial by Pincus et al.[46] the results at 6 weeks (before crossover) are introduced in the analysis. SMD = standardised mean difference, CI = confidence interval

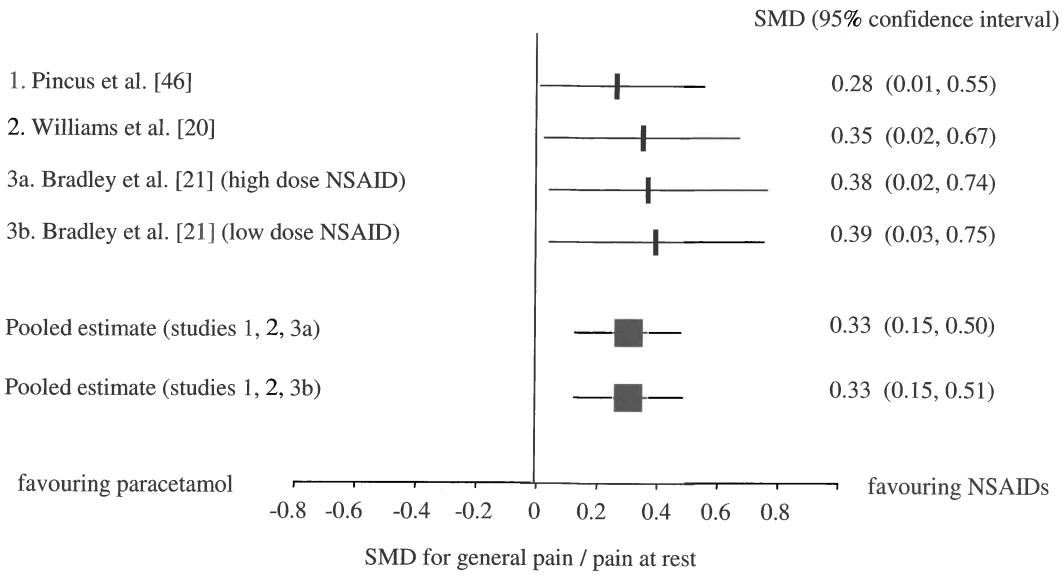

Figure 1. Differences between NSAIDs and paracetamol for pain at rest or general pain: results of individual studies and pooled standardised mean difference (SMD) at 4 to 6 weeks follow-up (random effects model, SMD and 95% confidence interval)

Appraisal of guidelines

Table 4 presents the results of the appraisal of the 9 selected guidelines using the AGREE criteria. Three guidelines satisfied more criteria than others, particularly regarding rigor of development.[5,19,24,51] The other guidelines did not sufficiently describe the methods used for searching and selecting evidence (item 8 and 9), nor did they define the methods used for formulating recommendations (item 10). Scope and purpose were adequately described in nearly all guidelines, but stakeholder involvement seemed to be limited in most guidelines: the patient's view and preferences (item 5), for example, were apparently only sought by the developers of the Canadian guideline [5] and the American College of Rheumatology (ACR) update [22], although the European League Against Rheumatism (EULAR) committee indicated the intent to do so.[24] A pilot among end-users (item 7) was only described for the Dutch, Canadian and North of England guidelines.[5,18,19] Most guidelines did not present information on potential organisational barriers of application of the recommendations (item 19), nor criteria for monitoring and auditing (item 21). Finally, editorial independence of the guideline committee (item 22) or conflicts of interest of individual members of guidelines committees (item 23) could not be established for the majority of the guidelines.

In Table 5 we summarise details of the guidelines regarding the target population, evidence base, and recommendations concerning the use of NSAIDs and paracetamol in OA. The recommendations of the German guideline were solely based on consensus among committee members.[23] The other guidelines mainly used results from RCTs, and included most of the evidence available at the time of development. However, none of the guidelines referred to the early crossover trials by Solomon and Abrams [49] and Wojtulewski, *et al* [50] identified by our search. These 2 studies may have been missed, or perhaps, were not considered to be of sufficient relevance or quality. Systematic reviews and meta-analyses were used [18,22] or carried out [5,19,24] by 5 guidelines. The ACR update also referred to patient preference studies, and in addition used evidence that had only been presented in abstracts or as a conference presentation.[22]

The actual content of recommendations regarding the use of NSAIDs or paracetamol was fairly consistent across most guidelines. The 3 guidelines that received relatively good scores for rigor of development all recommend paracetamol (maximum 4 g per day) as first-line therapy in OA.[5,19,24] NSAIDs are recommended for patients

Table 4. Results of the appraisal of guidelines for the management of osteoarthritis, using the AGREE criteria [34]

	BSR, 1993 [17]	ACR hip, 1995 [3]	ACR knee, 1995 [4]	Netherlands, 1998 [18]	North of England, 1998 [19]	ACR update, 2000 [22]	Canada, 2000 [5]	EULAR, 2000 [24]	Germany, 2000 [23]	Positive scores per item (%)
Scope and purpose										
1	+	+	+	+	+	+	+	+	+	9 (100%)
2	+	-	-	+	+	+	+	+	+	7 (78%)
3	+	+	+	+	+	+	+	+	+	9 (100%)
Stakeholder involvement										
4	+	?	?	-	+	?	+	+	+	5 (56%)
5	-	-	-	-	-	+	+	-	-	2 (22%)
6	?	-	-	+	+	-	+	-	+	4 (44%)
7	?	?	?	+	+	?	+	-	?	3 (33%)
Rigour of development										
8	-	?	?	?	+	?	+	+	-	3 (33%)
9	-	-	-	-	+	-	-	+	-	2 (22%)
10	-	-	-	?	+	?	+	+	-	3 (33%)
11	+	+	+	+	+	+	+	+	+	9 (100%)
12	-	-	-	-	+	?	+	+	-	3 (33%)
13	?	?	?	+	+	?	+	?	?	3 (33%)
14	-	-	-	+	+	-	-	+	+	4 (44%)
Clarity and presentation										
15	+	+	+	+	+	+	+	+	-	8 (89%)
16	+	+	+	-	+	+	+	+	+	8 (89%)
17	-	+	+	-	+	-	+	+	-	5 (56%)
18	?	?	?	+	?	?	?	?	?	1 (11%)
Applicability										
19	+	-	-	-	-	-	-	-	-	1 (11%)
20	-	+	-	-	+	+	+	+	-	5 (56%)
21	+	-	-	-	-	-	-	-	-	1 (11%)
Editorial independence										
22	?	?	?	?	+	+	+	?	?	3 (33%)
23	?	?	?	?	?	?	?	?	?	- (0 %)
Positive items (%)	9 (39%)	7 (30%)	6 (26%)	10 (43%)	18 (78%)	9 (39%)	17 (74%)	14 (61%)	8 (35%)	

BSR = British Society for Rheumatology, ACR = American College of Rheumatology, EULAR = European League Against Rheumatism

Table 5. Details of guidelines on the management of osteoarthritis (OA), including recommendations regarding prescription of NSAIDs or paracetamol

Guideline [reference]	Target population	Evidence base	Evidence used	Recommendation
BSR, 1993 [17]	OA hip and knee - clinical criteria - radiography not always needed	- no systematic evaluation of the literature - no explicit link between evidence and recommendations	RCTs: [21]	1. Analgesics (paracetamol). 2. NSAIDs are considered if adequate doses of paracetamol are ineffective: small dose, regularly re-assess need for NSAID therapy.
ACR hip, 1995 [3]	OA hip - clinical criteria and - radiography	- no clear presentation of a systematic evaluation of the literature - no explicit link between evidence and recommendations	RCTs: none available for hip OA. References for knee OA: [20,2139*]	1. Paracetamol (max. 4 g daily). 2. If inadequate, low-dose ibuprofen (< 1600 mg daily) or non-actylated salicylates. 3. If inadequate, full-dose NSAIDs.
ACR knee, 1995 [4]	OA knee - clinical criteria and/or - radiography	- no clear presentation of a systematic evaluation of the literature - no explicit link between evidence and recommendations	RCTs: [20,21,39*]	1. Paracetamol (max. 4 g daily). 2. If inadequate, low-dose ibuprofen (< 1600 mg daily) or non-actylated salicylates. 3. If inadequate, full-dose NSAIDs.
Netherlands, 1998 [18]	OA knee - clinical criteria	- systematic review of the literature mentioned, but not clearly presented - no explicit link between evidence and recommendations	RCTs: [20,21,47] plus systematic review	1. Paracetamol, 3-4 g daily
North of England, 1998 [19]	Joint pain believed to be caused by degenerative arthritis	- systematic review and meta-analysis, search strategies are available - explicit link between evidence and recommendations, using a grading system	RCTs: [20,21,43*]	1. Paracetamol (max. 4 g daily) 2. If inadequate, low-dose ibuprofen (1200 mg daily). 3. If inadequate, add paracetamol (max. 4 g daily), or high dose ibuprofen (2400 mg daily). 4. If inadequate, alternative drugs (diclofenac, naproxen or other NSAIDs, or co-codamol)

BSR = British Society for Rheumatology, ACR = American College of Rheumatology, EULAR = European League Against Rheumatism
* These RCTs were not included in our review, as they did not meet our selection criteria

Table 5. Continued

Guideline [reference]	Target population	Evidence base	Evidence used	Recommendation
ACR update, 2000 [22]	OA hip and knee - ACR criteria	- use of systematic reviews - strongest weight given to results of systematic reviews and RCTs	RCTs: [20,21] abstracts [45*,46] plus systematic reviews plus patient preference studies	1. Paracetamol (max. 4 g daily) 2. If inadequate: evaluation of risk factors for serious upper GI and renal toxicity, than COX-2-specific inhibitor or NSAIDS (plus misoprostol) 3. NSAIDs for patients with moderate to severe pain or signs of inflammation
Canada, 2000 [5]	OA - clinical criteria - radiography rarely needed	- extensive literature review, but search strategy not presented - explicit link between evidence and recommendations, using a grading system	RCTs: [20,21,39*]	1. Paracetamol (max. 4 g daily) for mild to moderate OA 2. NSAIDS may be used for moderate to severe OA, and in those whose symptoms are inadequately controlled by paracetamol
EULAR, 2000 [24]	OA knee - clinical criteria and/or - radiography	- systematic review of the literature - explicit link between evidence and recommendations, using a grading system	RCTs: [20,21,39*]	1. Treatment should be tailored to the individual patient 2. Paracetamol is the treatment of first choice 3. NSAIDs should be considered in those unresponsive to paracetamol (although there is no direct evidence base).
Germany, 2000 [23]	OA knee - clinical criteria and - radiography	- no systematic review of the literature - no explicit link between evidence and recommendations: consensus based.	none used	1. NSAIDs are the basis of therapy. 2. Low dose NSAIDs for those at relatively high risk of adverse events. Paracetamol can be used to reduce use of NSAIDs

BSR = British Society for Rheumatology, ACR = American College of Rheumatology, EULAR = European League Against Rheumatism
* These RCTs were not included in our review, as they did not meet our selection criteria

with moderate to severe symptoms, and for those who are unresponsive to paracetamol. The North of England guideline proposes a more detailed stepped-care approach for the pharmacological management of OA.[19] The EULAR guideline explicitly emphasises the need to tailor therapy to the individual patient.[24] Most guidelines recommend avoiding long-term use of NSAIDs in patients with a relatively high risk of adverse events, and some even present a risk profile.[3-5,22] In addition, 2 guidelines recommend regular reassessment of the need for NSAID therapy in patients who have been taking NSAIDs for some time.[17,22]

NSAIDs seem to be the mainstay of pharmacological treatment only in the German guideline, with paracetamol as an alternative in patients with a high risk of adverse events.[23] In contrast, the Dutch guideline, which was developed by general practitioners, states a clear preference for paracetamol. This guideline more strongly emphasises the increased risk of adverse events associated with NSAIDs, as well as the limited benefits of NSAIDs compared to paracetamol.[18]

Discussion

We conducted a systematic review of available evidence on the relative efficacy of NSAIDs compared to paracetamol for OA of the hip or knee, which showed consistent, yet modest, differences on pain in favour of NSAIDs (pooled SMD 0.33). The evidence base was relatively small. The search identified a large number of trials evaluating the efficacy of NSAIDs, but only few included a comparison with paracetamol. Our search strategy identified trials published until December 2001. Two relevant additional RCTs have been published since that time. Geba, *et al* [52] showed that the selective cyclooxygenase-2 inhibitor rofecoxib provides advantages over celecoxib and paracetamol in patients with knee OA. In a small study, Case, *et al* [53] reported larger clinical improvements in patients treated with diclofenac compared to paracetamol. However, differences in improvements between intervention groups were not statistically significant, possibly due to insufficient statistical power.

The important question is whether the magnitude of the differences between NSAIDs and paracetamol merits the initial use of NSAIDs. In the 2 most frequently cited

trials, the differences between NSAIDs and paracetamol at 4-6 weeks' follow-up were statistically significant only for pain at rest.[20,21] With the publication of the trial by Pincus, *et al* [46], the beneficial effects of NSAIDs seemed to be more strongly emphasised.[22,29]

Statistical significance should not dominate the discussion. It is more important to decide on the clinical relevance of the reported differences. This is difficult and certainly arbitrary, as it will depend on several factors, including the severity of the condition, potential side effects, inconvenience of therapy, treatment preferences, and costs.[54] For 50-foot walk time the difference in improvement between NSAIDs and paracetamol was less than 0.7 second [20,21], a difference that will be considered relevant by few. For general pain or pain at rest the difference in improvement was roughly 8 points on a visual analogue scale (range 0 to 100),[46] 0.8 point on a numeric rating scale (0-10),[20] or 0.3 point on an ordinal 4-point scale [21]. Taking the mean baseline score into account the average difference in response was 17 to 27% in favour of NSAIDs. This seems to be a relevant treatment effect; research in patients with OA has shown that a difference of about 15% on pain or 30% on function is associated with other predefined definitions of clinically important change.[55,56] The question remains whether this difference is large enough to recommend the use of NSAIDs, despite the higher risk of adverse reactions.

There appears to be wide variation among patients in the response to NSAIDs. Scott-Lennox, *et al* [57] suggested that the washout period that often precedes randomisation in clinical trials may affect the response to therapy. In their study, patients with higher intensity flares during washout were more likely to report substantial improvement of symptoms, regardless of treatment (active or placebo). Most trials selected for our review included a washout period before baseline assessment. Little information is available about the course of symptoms during washout (flare intensity), but in 2 of the larger trials pain scores increased during washout.[21,46] Secondary, post hoc analyses of the trial by Bradley, *et al* [21] seem to confirm the assumption of Scott-Lennox, *et al*: a larger improvement of pain was found for patients with higher pain levels after washout, with no significant differences between intervention groups.[48] On the other hand, subgroup analyses by Pincus, *et al* [46] showed that patients with higher pain scores after washout seemed to profit more from NSAIDs compared to paracetamol.

The ACR update [22] recommends NSAIDs as initial therapy for patients with more severe pain, and also for those with signs of inflammation. Secondary analyses [47] of the trial by Bradley, *et al* [21] did not provide evidence for a stronger effect of NSAIDs in patients with signs of inflammation, such as joint tenderness or swelling. Additional research is needed to explore and confirm relevant subgroup differences. This requires randomised trials in sufficiently large populations using prestratification based on baseline severity of pain, function, or other indicators of disease severity. Such research may facilitate the early identification of patients for whom paracetamol will suffice, and those who benefit more from NSAIDs.

In addition to treatment efficacy, drug preferences are the result of several factors, including subjective benefit, side effects (actual or potential), ease of administration (e.g., the larger number of tablets per day when using paracetamol may be perceived as bothersome), doctor or patient beliefs and interactions, and severity of disease.[25,58] Preferences of patients for either NSAIDs or paracetamol have been assessed in a large survey [25] and in a few randomised trials.[46,49] About 60% of patients in these studies preferred NSAIDs to paracetamol. However, this also means that paracetamol may be satisfactory to a considerable proportion of patients. In the trial by Pincus, *et al* 42% of subjects and care providers, while still blinded to treatment allocation, rated paracetamol as better than or equally effective compared to NSAIDs.[46] Therefore, considering the higher risk of adverse reactions to NSAIDs, it seems worthwhile to recommend continued use of paracetamol in these patients as long as symptoms do not greatly interfere with daily activities. Paracetamol may also be recommended in patients who report previous use of paracetamol, but have not tried adequate doses.

The appraisal of guidelines showed that the methods for developing the guideline were often unclear. Additional information on the development process was not available for all guidelines. The addition of a category for scoring absence of information facilitated the appraisal, and may have prevented overly negatives scores for some guidelines. Nonetheless, guideline development often appeared to be inadequate, particularly regarding methods for selecting evidence and formulating recommendations. Clear instructions for developing guidelines have become available,[59,60] and some of the recent guidelines clearly meet these higher standards.[5,19,24] Further improvements can be made in the description of stakeholder involvement, applicability, and editorial

independence. The fact that most guidelines do not clearly state that there was no conflict of interest for individual members of the guideline committee was surprising. We argue that editorial independence is of utmost importance, particularly for guidelines in which the pharmaceutical industry may have a strong interest.

In the EULAR guidelines, the authors indicate that there was often discordance between research evidence and the opinion of experts.[24] In this international guideline, variation across countries regarding health care delivery systems and attitudes toward OA contributed to this discordance. A Delphi system was used to obtain consensus on difficult issues. Despite variation in the design of the guidelines in our review, the actual content of recommendations regarding the use of NSAIDs or paracetamol in OA was fairly consistent, with most guidelines recommending paracetamol (maximum 4 g per day) as the first-line therapy in OA. The results of a nested case-control study [61] and a recent retrospective cohort study [62] indicate that patients who take high dose paracetamol may be more likely to experience GI events compared with those taking low dose paracetamol, and that these risks may be similar to high dose NSAIDs. Both studies controlled for known risk factors of confounding by indication, i.e., the risk that patients who are more likely to suffer GI complications are more likely to be prescribed paracetamol than NSAIDs. However, in observational research it is almost impossible to completely avoid or adjust for confounding by indication.[61] RCTs with long-term follow-up are warranted to prospectively compare the benefits and risks of high dose paracetamol with high dose NSAIDs.

Guidelines are designed to assist clinicians in making decisions about appropriate health care in specific situations. Guidelines generally deal with the 'average' patient, and may not always be appropriate for individual patients, each of whom may have different needs and expectations. Several factors need to be taken into account that may vary considerably across patients, such as co-morbidity, co-medication, daily activity requirements, patient preferences and the patient' s perceptions and knowledge of OA.[63] The need to tailor management to individual patients was explicitly stated in the EULAR guidelines.[24] In an editorial, Dieppe argued that guidelines are too constrained to be useful in the management of a chronic heterogeneous condition such as OA. A better alternative may be to lay out the options available, with information on the advantages and disadvantages of each option that can be understood by both patients and health care

providers.[64] This can be a welcome addition to current guidelines, particularly in patients for whom the expected benefits of NSAIDs may be limited (e.g., patients with mild symptoms), and may also facilitate processes of shared decision-making between patient and physician.[62,63] Further, N-of-1 trials may be helpful to resolve the question for individual patients whether paracetamol is equally as effective as NSAIDs in the management of pain and disability.[67,68]

In conclusion, evidence from RCTs shows that the benefits of NSAIDs on pain are significantly greater than those of paracetamol, but the difference was small and may be of limited value to a large proportion of patients with OA. This supports the recommendations of recent guidelines to use paracetamol as initial therapy for OA (in addition to non-pharmacological interventions). Further research is needed to establish the efficacy of NSAIDs or paracetamol in relevant subgroups of patients. Nevertheless, we agree with recently developed guidelines that it is important that treatment is tailored to individual patients, taking into account the severity of symptoms, risk of side effects, previous use of paracetamol, and patient's expectations and preferences.

Acknowledgements

We gratefully acknowledge professor M. Boers, MD, PhD (rheumatology specialist) for his critical review of our paper and useful suggestions. We thank all the general practitioners and their patients who participated in the study. This study was supported unconditionally by Leo Pharma, The Netherlands.

References

1. Lawrence RC, Helmick CG, Arnett FC, Deyo RA, Felson DT, Giannini EH et al. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthrit Rheum 1998;41:778-99.
2. Van Saase JL, van Romunde LK, Cats A, Vandenbroucke JP, Valkenburg HA. Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations. Ann Rheum Dis 1989;48:271-80.
3. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR et al. Guidelines for the medical management of osteoarthritis. Part I. Osteoarthritis of the hip. American College of Rheumatology. Arthrit Rheum 1995;38:1535-40.
4. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR et al. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. American College of Rheumatology. Arthrit Rheum 1995;38:1541-6.
5. Holbrook AM (Chair) for Ontario Musculoskeletal Therapy Review Panel. Medical treatment guidelines for the treatment of osteoarthritis, rheumatoid arthritis, and acute musculoskeletal injury. Toronto: Ontario Ministry of Health and Long-term Care, 2000.
6. Bierma-Zeinstra SM, Lipschart S, Njoo KH, Bernsen R, Verhaar J, Prins A et al. How do general practitioners manage hip problems in adults? Scand J Prim Health Care 2000;18:159-64.
7. Traynor V, Britt H, Neary S, Sayer GP, Charles J, Meza RA. The management of osteoarthritis in general practice. Results from the Australian Morbidity and Treatment Survey, 1990-1991. Aust Fam Phys 1994;23:1971-8.
8. Mamlin LA, Melfi CA, Parchman ML, Guttierez B, Allen DL, Katz BP et al. Management of osteoarthritis of the knee by primary care physicians. Arch Fam Med 1998;7:563-7.
9. Mazzucca SA, Brandt KD, Katz BP, Dittus RS, Freund DA, Lubitz R et al. Comparison of general internists, family physicians, and rheumatologists managing patients with symptoms of osteoarthritis of the knee. Arthr Care Res 1997;10:289-99.
10. Griffin M, Piper JM, Daugherty JR, Snowden M, Ray WA. Nonsteroidal anti-inflammatory drug use and increased risk of peptic ulcer disease in elderly persons. Ann Int Med 991;114:257-63.

11. Singh G, Rosen Ramey D. NSAID induced gastrointestinal complications: the ARAMIS perspective - 1997. Arthritis, Rheumatism, and Aging Medical Infromation System. J Rheum Suppl 1998;51:8-16.

12. Henry D, Lim LL, Garcia Rodriguez LA, Perez Gutthann S, Carson JL, Griffin M et al. Variability in the risk of gastrointestinal complications with individual non- steroidal anti-inflammatory drugs: results of a collaborative meta-analysis. BMJ 1996;312:1563-6.

13. Fored CM, Ejerblad E, Lindblad P, Fryzek JP, Dickman PW, Signorello, LB et al. Acetaminophen, aspirin, and chronic renal failure. N Engl J Med 2001;345:1801-1808.

14. Whitcomb DC, Block GD. Association of acteminophen hepatotoxicity with fasting and ethanol use. JAMA 1994;272:1845-50.

15. Schiodt FV, Rochling FA, Casey DL, Lee WM. Acetaminophen toxicity in an urban county hospital. N Engl J Med 1997;337:1112-7.

16. Hylek EM, Heiman H, Skates SJ, Sheehan MA, Singer DE. Acetaminophen and other risk factors for excessive warfarin anticoagulation. JAMA 1998;279:657-62.

17. Scott DL. Guidelines for the diagnosis, investigation and management of osteoarthritis of the hip and knee. Report of a Joint Working Group of the British Society for Rheumatology and the Research Unit of the Royal College of Physicians. J R Coll Physicians Lond 1993;27:391-6.

18. Bijl D, Dirven-Meijer PC, Opstelten W, Raaijmakers AJ, Scholten RJPM, Eizinga WH et al. General practice guidelines for non-traumatic knee complaints in adults. [in Dutch: NHG-standaard Niet-traumatische Knieproblemen bij volwassenen.] Huisarts Wet 1998;41:344-50.

19. Eccles M, Freemantle N, Mason J. North of England evidence based guideline development project: summary guideline for non-steroidal anti-inflammatory drugs versus basic analgesia in treating the pain of degenerative arthritis. BMJ 1998;317:526-30.

20. Williams HJ, Ward JR, Egger MJ, Neuner R, Brooks RH, Clegg DO et al. Comparison of naproxen and acetaminophen in a two-year study of treatment of osteoarthritis of the knee. Arthritis Rheum 1993;36:1196-206.

21. Bradley JD, Brandt KD, Katz BP, Kalasinki LA, Ryan SI. Comparison of an antiinflammatory dose of ibuprofen, an analgesic dose of ibuprofen, and acetaminophen in the treatment of patients with osteoarthritis of the knee. N Engl J Med 1991;325:87-91.

22. American College of Rheumatology Subcommittee on osteoarthritis guidelines. Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. Arthrit Rheum 2000;43:1905-15.

23. Puhl W, Bernau A, Böhle E. Brune K, Gerhardt P, Greitemann B et al. Ambulatory diagnosis and therapy of gonarthrosis. [in German: Ambulante diagnostik und Therapie der Gonarthrose.] Z Orthop 2000;138:85-93.

24. Pendleton A, Arden N, Dougados M, Doherty M, Bannwarth B, Bijlsma JW et al. EULAR recommendations for the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis 2000;59:936-44.

25. Wolfe F, Zhao S, Lane N. Preference for nonsteroidal antiinflammatory drugs over acetaminophen by rheumatic disease patients. Arthrit Rheum 2000;43:378-85.

26. Moskowitz RW. The role of anti-inflammatory drugs in the treatment of osteoarthritis: a United States viewpoint. Clin Exp Rheumatol 2001;19 Suppl 25:S3-8.

27. Hungin APS, Kean WF. Non-steroidal anti-inflammatory drugs: overused or underused in osteoarthritis? Am J Med 2001;110:8S-11S.

28. Jawad AS. EULAR recommendations for the management of knee osteoarthritis. Ann Rheum Dis 2001;60:540.

29. Felson DT. The verdict favors nonsteroidal antiinflammatory drugs for treatment of osteoarthritis and a plea for more evidence on other treatments. Arthrit Rheum 2001;44:1477-80.

30. Dieppe PA. Concerns about the methodology used in developing the 2000 update of the American College of Rheumatology recommendations for management of hip and knee osteoarthritis.[letter] Arthrit Rheum 2001;44:2450-1.

31. Brandt KD. A critique of the 2000 update of the American College of Rheumatology recommendations for management of osteoarthritis.[letter] Arthrit Rheum 2001;44:2451-5.

32. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. BMJ 1994;309:1286-91.

33. Van Tulder MW, Scholten RJ, Koes BW, Deyo RA. Nonsteroidal anti-inflammatory drugs for low back pain: a systematic review within the framework of the Cochrane Collaboration Back Review Group. Spine 2000;25:2501-13.

34. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. New Engl J Med 1988;318:1728-33.

35. Cohen J. Statistical power analysis for the behavioral sciences [2nd Ed]. Hills Dale, New Jersey: Lawrence Erlbaum Associates, 1988.

36. DerSimonian R, Laird N. Meta-analysis in clinical trials. Contr Clin Trials 1986;7:177-88.

37. Fleiss JL. The statistical basis of meta-analysis. Stat Methods Med Res 1993;2:121- 45.

38. The AGREE Collaboration. Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument. www.agreecollaboration.org.

39. Amadio P, Cummings DM. Evaluation of acetaminophen in the management of osteoarthritis of the knee. Curr Ther Res 1983;34:59-66.

40. Amor B. Benarrosh C. A method for comparing anagesics: glafenine and paracetamol multicenter cross-over approach. Clin Rheumatol 1988;7:492-7.

41. Finkbeiner G. Treatment of the diseases of the extended rheumatic sphere form with paracetamol. [in German: Behandlung von Erkrankungen des erweiterten rheumatischen Formenkreises mit Paracetamol (Dolo-Arthrosetten).] Münch Med Wochenschr 1973;115:1855-7.

42. Kjaersgaard-Andersen P. Evaluating codeine plus paracetamol for pain. Nursing Times 1991;87:52.

43. Parr G, Darekar B, Fletcher A, Bulpitt CJ. Joint pain and quality of life; results of a randomised trial. Br J Clin Pharmacol 1989;27:235-42.

44. Brooks PM, Dougan MA, Mugford S, Meffin E. Comparative effectiveness of 5 analgesics in patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 1982;9:723-6.

45. Altman RD for the IAP Study Group. Ibuprofen, acetaminophen, and placebo in osteoarthritis of the knee: a six-day double-blind study.[Abstract] Arthrit Rheum;1999:42 Suppl 9:S403.

46. Pincus T, Koch CG, Sokka T, Lefkowitz J, Wolfe F, Jordan JM et al. A randomized double-blind, crossover clinical trial of diclofenac plus misoprostol versus acetaminophen in patients with osteoarthritis of the hip or knee. Arthrit Rheum 2001;44:1587-98.

47. Bradley JD, Brandt KD, Katz BP, Kalasinski LA, Ryan SI. Treatment of knee osteoarthritis: relationship of clinical features of joint inflammation to the response to a nonsteroidal antiinflammatory drug or pure analgesic. J Rheumatol 1992;19:1950-4.

48. Bradley JD, Katz BP, Brandt KD. Severity of knee pain does not predict a better response to an antiinflammatory dose of ibuprofen than to analgesic therapy in patients with osteoarthritis. J Rheumatol 2001;28:1073-6.

49. Solomon L, Abrams G. Orudis in the management of osteoarthritis of the knee. A double-blind trial. S Afr Med J 1974;48:1526-9.

50. Wojtulewski JA, Hart FD, Huskisson EC. Fenoprofen in treatment of osteoarthritis of the hip and knee. BMJ 1974;2:475-6.

51. Eccles M, Freemantle N, Mason J. North of England evidence based guidelines development project: methods of developing guidelines for efficient drug use in primary care. BMJ 1998;316:1232-5.

52. Geba GP, Weaver AL, Polis AB, Dixon ME, Schnitzer TJ. Efficacy of rofecoxib, celecoxib, and acetaminophen in osteoarthritis of the knee – a randomized trial. JAMA 2002;287:64-71.

53. Case JP, Baliunas AJ, Block JA. Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis. Arch Intern Med 2003;163:169-178.

54. Cook DJ, Guyatt GH, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest 1992;102:305S-11S.

55. Dougados M, LeClaire P, Van der Heijde D, Bloch DA, Bellamy N, Altman RD. Response criteria for clinical trials on osteoarthritis of the knee and hip. Osteoarthritis and Cartilage 2000;8:395-403.

56. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Univeristies Osteoarthritis index questionnaire and global assessment in patients with osteoarthritis. J Rheum 2000;27:2635-41.

57. Scott-Lennox JA, McLaughlin-Miley C, Lennox RD, Bohlig AM, Cutler BL, Yan C, Jaffe M. Stratification of flare intensity identifies placebo responders in a treatment efficacy trial of patients with osteoarthritis. Arthrit Rheum 2001;44:1599-1607.

58. Pincus T, Swaeringen C, Cummings P, Callahan LF. Preference for non-steroidal antiinflammatory drugs versus acetaminophen and concomitant use of both types of drugs in patients with osteoarthritis. J Rheumatol 2000;27:1020-7.

59. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Developing guidelines. BMJ 1999;318:593-6.

60. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ 2001;323:334-6.

61. Garcia Rodriguez LA, Hernandez-Diaz S. The risk of upper gastrointestinal complications associated with nonsteroidal anti-inflammatory drugs, glucocorticoids, acetaminophen, and combinations of these agents. Arthritis Res 2001;3:98-101.

62. Rahme E, Pettitt D, LeLorier J. Determinants and sequelae associated with utilization of acetaminophen versus traditional nonsteroidal antiinflammatory drugs in an elderly population. Arthrit Rheum 2002;46:3046-54.

63. Courtney P, Doherty M. Key questions concerning acetaminophen and NSAIDs for osteoarthritis. Ann Rheum Dis 2002;61:767-73.

64. Dieppe P. From protocols to principles, from guidelines to toolboxes: aids to good management of osteoarthritis. Rheumatology 2001;40:841-2.

65. Edwards A, Elwyn G.Understanding risk and lessons for clinical risk communication about treatment preferences. Qual Health Care 2001;10 Suppl 1:9-13.

66. Elwyn G, Edwards A, Kinnersley P, Grol R. Shared decision making and the concept of equipoise: the competences of involving patients in healthcare choices. Br J Gen Pract 2000;50:892-9.

67. March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. N of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. BMJ 1994;309:1041-5.

68. Nikles CJ, Glasziou PP, Del Mar CB, Duggan CM, Clavarino A, Yelland MJ. Preliminary experiences with a single-patient trials service in general practice. Med J Aust 2000;173:100-3.

Abstract

Objectives: To investigate for individual patients who have been using non-steroidal anti-inflammatory drugs (NSAIDs) regularly, whether paracetamol is as effective as NSAIDs in the treatment of pain and disability related to osteoarthritis of the hip or knee.

Methods: A series of N-of-1 trials were conducted in general practices in Amsterdam and the surrounding area. Each patient was his or her own control and received five pairs of treatments comprising two weeks of an NSAID and two weeks of paracetamol. For each pair, the order of treatments was randomised. Outcome measures were severity of individual main complaints, intensity of pain, satisfaction with drugs, function test, and side effects.

Results: Thirteen patients were selected. Six patients did not complete the study. For five patients completing the study little or no difference was found between NSAIDs and paracetamol, for one patient the results favoured the NSAID, and for one patient there was no association between outcome and type of drug. It was recommended that six patients changed to paracetamol; the others continued with NSAIDs. Three months after the end of the study, four of the six patients for whom paracetamol had been recommended were taking NSAIDs for practical reasons or perceived lack of efficacy.

Conclusions: The results of the N-of-1 trials varied across patients. N-of-1 trials can be used to investigate which treatment is best for any specific person, thus avoiding unnecessary prolonged treatment with NSAIDs. However, practical reasons may cause patients to switch from NSAIDs to paracetamol or not.

Introduction

Osteoarthritis (OA) is a common musculoskeletal disorder. The prevalence of OA increases with age to over 50% in the knee and about 20% in the hip in women aged 80 years and older. In elderly men the prevalence of OA of the knee is about 25%, and the hip is affected in more than 10%.[1]

Non-steroidal anti-inflammatory drugs (NSAIDs) are frequently prescribed for OA. Available evidence from three randomised trials on the relative effectiveness of NSAIDs compared with paracetamol (acetaminophen) for OA of the hip or knee shows consistent, yet modest, differences in their effect on pain in favour of NSAIDs.[2-6] However, it is generally known that NSAIDs are associated with gastrointestinal complications, whereas paracetamol is associated with very few side effects.[7-9] Therefore, the national practice guidelines for OA management, issued by the Dutch College of General Practitioners, recommend paracetamol as initial treatment in the treatment of pain and disability related to OA of the knee.[10] Similar recommendations are made in the EULAR guidelines and the British, Canadian and American guidelines.[11-14]

The EULAR guideline explicitly emphasises the need to tailor treatment to the individual patient. Indeed, for OA there may be reasonable doubt whether the results of large randomised controlled trials (RCTs) can be generalised to individual patients in general practice. For instance, a patient who consults a general practitioner for chronic knee pain and stiffness may not meet all the selection criteria of the RCTs, reported outcome measures may not be relevant for this specific patient, or the type or dosage of drugs studied in the RCTs may not be similar to the drugs the patient is using. Normally, when a doctor has any doubts about the applicability of treatment recommendations to a specific patient, the trial and error method is used. This means that a particular drug will be prescribed, continued if considered effective, or changed if considered not beneficial. However, the fact that both patient and doctor are aware of the change in treatment can influence the evaluation of its effectiveness. Individual preferences and expectations of both doctor and patient can play an important part in the assessment. If the doctor or the patient doubts the effectiveness of a treatment with a drug, and prefers an objective evaluation, the N-of-1 trial is a useful tool.[15]

N-of-1 trials are carried out with one patient. The patient is his or her own control and receives several periods of both the intervention treatment and the control treatment. The sequence of treatments is randomised. Because N-of-1 trials are meant to evaluate the effects of treatments for each individual patient (and not to estimate the average effect for a larger population), the size of a series of N-of-1 trials is of minor importance. Nevertheless, a series of N-of-1 trials will be able to demonstrate the potential variation in outcome across patients.

Knowing that patients with OA may continue to use analgesics for a prolonged period of time, it is worthwhile reconsidering drug treatment. This study aimed at investigating for individual patients who have been using NSAIDs regularly, whether paracetamol is as effective as NSAIDs in the treatment of pain and disability related to OA of the hip or knee.

Methods

Patient selection

Twenty-five general practitioners in Amsterdam and the surrounding area were asked to select patients from their medical records. Selection criteria were OA of the hip or knee (as diagnosed by the general practitioner); regular use of diclofenac, ibuprofen, or naproxen (at least for a period of one month, five days a week); no contraindications for NSAIDs; no contraindications for paracetamol; no corticosteroid injections in the three months preceding the study; no (planned) arthroplasty; and no concomitant physiotherapeutic treatment. Additionally, the patients had to be able to fill in a questionnaire in Dutch. If more than one hip or knee was affected by OA the patient was asked to indicate the joint that caused the most severe complaints.

Design

The study comprised a series of N-of-1 trials. Each patient received five pairs of treatments starting between November 1999 and May 2000. Each pair consisted of a two-week period of NSAID treatment and a two-week period of paracetamol treatment. For

each pair of treatments the order of NSAIDs and paracetamol was randomised. The random sequence was prepared in advance by the hospital pharmacist for each patient separately. The patient, the general practitioner, and the investigator were unaware of the sequence of treatment. Identical capsules with similar colour, smell, and taste were used to ensure blinding. If necessary, placebos were given, so that the patients received the same number of capsules in all treatment periods. For example, if 1000 mg paracetamol three times daily was compared with 50 mg diclofenac twice daily, the patient received additional placebos once daily in each diclofenac period. The study was approved by the ethics review board of the VU University Medical Center, and all participating patients gave written informed consent at the start of the study.

Drug treatment

During the NSAID treatment periods each patient received the same type of NSAID and, if possible, in the same dosage, as before the start of the study. The dosage of paracetamol was adjusted to the dosage of the NSAID. For example, if a patient took 400 mg ibuprofen three times daily, 1000 mg paracetamol was prescribed three times daily. Maximum dosages were 3000 mg paracetamol divided into three or four doses a day, 400 mg ibuprofen four times daily, 50 mg diclofenac twice daily or 25 mg diclofenac three times daily, and 750 mg naproxen divided into two doses a day. Pill counts were taken to measure compliance.

Outcome assessment

Outcomes were assessed by daily diaries and during home visits made by the investigator every two weeks at the end of each treatment period. Because of the possibility of carry-over effects in the first week of each treatment period, only data for the second week of each treatment period were taken into account in the analysis. We considered a washout period of one week sufficient, given the short half life of the NSAIDs used in our study.[9] Previous N-of-1 series and RCTs comparing NSAIDs with paracetamol have used a similar washout period.[2,6,16-18]

Primary outcome measures were *(a)* individual main complaints and *(b)* the intensity of pain. To compose a total diary score, each patient was instructed during the selection procedure to identify their individual main complaints of pain, stiffness, or

limitations in daily functioning. These individual main complaints (four to eight symptoms for each patient) were scored daily on seven point ordinal scales ranging from 0 = no complaints at all to 6 = unbearable complaints.[19,20] A total diary score was computed by calculating for each day the median score of all these four to eight individual main complaints. The severity of the most important individual complaint, according to the patient, was analysed separately. The intensity of pain during the previous week was scored by the patient on an 11-point numerical rating scale (0-10, 10 indicating very severe pain) after each treatment period during the home visits made by the investigator.

Secondary outcome measures were *(a)* satisfaction with drugs, scored on a seven point ordinal scale ranging from 0 = very satisfied to 6 = very dissatisfied and *(b)* one of three possible function tests partially derived from previously described function tests [21] (walking 2.5 metres to and fro, sitting down and standing up five times, or going up and down stairs [time in seconds]). Every two weeks the same function test was performed by the patient. The choice of the type of test depended on the main complaints of the patient, the ability of the patient to perform the function test, and the possibility that the test could be performed at the patient' s home. Both secondary outcome measures were scored every two weeks. Furthermore, every two weeks, information was collected on the occurrence of side effects and co-interventions (including concomitant drugs and other conservative methods of treatments).

The influence of the results of the N-of-1 trials on the treatment policy of the general practitioners was also investigated. Both before the start of the study and at the end of the study each general practitioner was asked which drug (NSAID, paracetamol or other) they thought to be best for each specific patient. Subsequently, the general practitioners scored how sure they were about this answer on a five point ordinal scale ranging from 1 = very certain to 5 = very uncertain.

Breaking the randomisation code

Before analysis of the data the randomisation code was partially broken. This means that the type of treatment was indicated as either A or B, without the investigator (AW) knowing whether A or B represented treatment with an NSAID or treatment with paracetamol. Subsequently, recommendations were formulated for further treatment.

After analysing and formulating the recommendations, the supervisor (DW) decoded the randomisation completely and conferred with the general practitioner about the results of the trial. Finally, the recommendation was discussed with the patient.

Analysis

Because the single case design was used, all outcome measures were analysed for each patient separately. For the second week of each treatment period, median scores were calculated for the total diary score and the individual main complaint. Differences in scores between the NSAID and paracetamol were calculated for each outcome measure. Finally, the proportion of pairs of treatment periods for which treatment A was better than treatment B, and vice versa, was calculated for the primary outcome measures.

Decisions for future treatment (recommendations)

The following rules were used to decide on treatment recommendations: firstly, the median difference between treatments for the total diary score had to be at least one point in favour of the NSAID to recommend NSAID treatment, otherwise paracetamol was recommended; secondly, the results for at least 75% of the treatment pairs had to be in favour of the NSAID for at least one of the primary outcome measures (for example, 4/5 pairs) to recommend NSAID treatment, otherwise paracetamol was recommended; and finally, improvements in severe complaints were considered to be more important than improvements in only mild or moderate complaints. Three months after completion of the N-of-1 trials, the patients were contacted by telephone and asked to indicate which drugs they were currently using for their OA.

Results

Subjects

A search in the medical records identified 22 patients who were eligible for participation according to their general practitioners. Nine patients were excluded for the following reasons: *(a)* no actual use of diclofenac, ibuprofen, or naproxen (three patients used paracetamol and one patient piroxicam); *(b)* no frequent use of NSAIDs (one patient);

(c) contraindications for NSAIDs (three patients); and *(d)* no informed consent (one patient used suppositories and refused oral drugs). Thirteen patients were selected for participation: 11 women and two men (Table 1). Their ages ranged between 50 and 91 years (median 77 years). In 12 patients the main problems concerned the knee, and in one patient the hip. Before the start of the study nine patients were taking diclofenac, three ibuprofen, and one naproxen.

Six patients did not complete the study: four because of perceived lack of efficacy, one because of perceived side effects, and one because of concomitant disease and loss of motivation (Table 1). After withdrawal, these patients continued to use NSAIDs. Seven patients did complete their 20-week trial period. All seven patients who completed the study took more than 94% of their trial capsules. From the third week, patient No 7 was prescribed additional paracetamol by the general practitioner for concomitant disease. As a consequence, for this patient the first two weeks could not be compared with the other 18 weeks, and only four pairs of treatments could be analysed.

Effectiveness of paracetamol compared with NSAID

In five patients who completed the study there was little or no difference in outcome between NSAIDs and paracetamol, and paracetamol was recommended (patients Nos 1, 3, 4, 5, and 9). In two patients (patients Nos 6 and 7) the median difference between treatments for the total diary score was one point in favour of NSAIDs (Table 2). The percentage of pairs in which NSAIDs were favoured above paracetamol was at least 75% for at least one of the primary outcome measures in both these patients. However, the improvements in complaints for patient No 7 mainly concerned changes from mild to very mild, whereas in patient No 6 improvements in complaints in favour of the NSAID were from severe to moderate. For these reasons, only patient No 6 was recommended to continue with NSAID treatment. For the other patients who completed the study (including patient No 7) paracetamol treatment was recommended.

Side effects and co-interventions

There were no differences in the occurrence of side effects between the NSAID and the paracetamol treatment periods. The patient who did not complete the study because of

Table 1. Patient characteristics and reasons for withdrawal of 13 participants with osteoarthritis of the hip or knee in a series of N-of-1 trials

Patient No	Gender (M/F)[a], age, joint	Drug treatment before the start of the trial	Trial drug	Number of weeks completed, Reasons for withdrawal
1	F, 71, hip	3 x day 400 mg ibuprofen	3 x day 400 mg ibuprofen versus 3 x day 1000 mg paracetamol	20
2	M, 82, knee	3 x day 50 mg diclofenac	2 x day 50 mg diclofenac versus 3 x day 1000 mg paracetamol	0, (only 3 days): perceived lack of efficacy (stiffness)
3	F, 80, knee	3 (sometimes 4) x day 400 mg ibuprofen	3 x day 400 mg ibuprofen versus 3 x day 1000 mg paracetamol	20
4	F, 91, knee	Mostly 1 x day 50 mg diclofenac (sometimes more)	1 x day 50 mg diclofenac versus 1 x day 1000 mg paracetamol and 1 x day 500 mg paracetamol	20
5	M, 50, knee	1 x day 50 mg diclofenac	1 x day 50 mg diclofenac versus 1 x day 1000 mg paracetamol and 1 x day 500 mg paracetamol	20
6	F, 54, knee	1 or 2 x day 500 mg naproxen	1 x day 250 mg naproxen and 1 x day 500 mg naproxen versus 3 x day 1000 mg paracetamol	20
7	F, 73, knee	1 x day 50 mg diclofenac	1 x day 50 mg diclofenac versus 1 x day 1000 mg paracetamol and 1 x day 500 mg paracetamol. After 2 weeks she was prescribed 1000 mg additional paracetamol, because of concomitant disease (gout).	20
8	F, 62, knee	1 or 2 x day 100 mg diclofenac	2 x day 50 mg diclofenac versus 2 x day 1000 mg paracetamol[b]	4: Perceived lack of efficacy (pain)
9	F, 82, knee	1 x day 75 mg diclofenac and 500 mg paracetamol	1 x day 25 mg diclofenac and 1 x day 50 mg diclofenac versus 3 x day 1000 mg paracetamol (after 4 weeks 2 x day 1000 mg paracetamol)	20
10	F, 72, knee	1 x day 50 mg diclofenac, sometimes 2 x day 50 mg diclofenac	1 x day 50 mg diclofenac versus 1 x day 1000 mg paracetamol and 1 x day 500 mg paracetamol	8: Concomitant disease (back pain) and loss of motivation
11	F, 77, knee	3 x day 50 mg diclofenac	2 x day 50 mg diclofenac versus 3 x day 1000 mg paracetamol	4: Perceived side effects (stomach complaints and nausea)
12	F, 79, knee	3 x day 50 mg diclofenac	2 x day 50 mg diclofenac versus 3 x day 1000 mg paracetamol	4:Perceived lack of efficacy (pain)
13	F, 90, knee	3 x day 400 mg ibuprofen (and almost every day 500 mg paracetamol/codeine, sometimes also 1 x day 500 mg paracetamol)	3 x day 400 mg ibuprofen versus 3 x day 1000 mg paracetamol	2: Perceived lack of efficacy (pain)

[a]M = male, F = female
[b]The patient received 2000 mg instead of 3000 mg paracetamol daily, because of alcohol use (approximately 6 units a day).

Table 2. Results of the seven patients who completed the N-of-1 trials: outcome measures, recommendations, medication after three months and confidence

Patient No	Joint	Most important complaint	Function test	Median difference N – P[a] (median N, median P)[b]					Proportion of pairs N>P for the primary outcome measures			Recom-mendation	Drug treatment after 3 months	δ confidence GP (choice before, choice after)[c]
				Total diary score	Individual main complaint	Pain	Satisfaction	Function test (seconds)	Total diary score	Individual main complaint	Pain			
1	H	pain during standing up from a chair or couch	W	0 (3,3)	0 (3,3)	-1 (6, 7)	0 (2, 2)	0.5 (10.5, 10.6)	1/5	1/5	3/5	P	N	0 (N, P)
3	K	pain during walking	W	-0.5 (3, 3.5)	0 (4, 4)	0 (6, 6)	0 (3, 3)	0 (12.0, 11.7)	3/5	0/5	0/5	P	N and P	1 (N, P)
4	K	pain during walking	W	0 (2.5, 2.5)	0 (3, 3)	0 (7, 7)	0 (3, 2)	-0.7 (16.0, 16.9)	2/5	2/5	2/5	P	N	0 (N, P)
5	K	pain during sleeping	W	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0.2 (6.8, 6.6)	0/5	0/5	2/5	P	P	2 (N, P)
6	K	pain during standing up from the toilet	ST	-1 (3, 4)	-1 (3, 4)	0 (7, 8)	0 (3, 5)	-0.8; -2.0 (12.2, 13.2; 17.7, 19.4)[d]	4/5	4/5	2/5	N	N	2 (N, N)
7[e]	K	pain during standing up from a chair	S	-1 (1, 2)	-0.5 (1, 2)	-3 (3, 5.5)	0.5 (1.5, 1)	0.8 (18.6, 18.9)	3/4	2/4	3/4	P	N	1 (P, P)
9	K	pain during standing	W	0 (3, 3)	0 (3, 3)	-1 (6, 7)	0 (2, 2)	-3.6 (20.2, 22.7)	1/5	1/5	3/5	P	N (and P)	0 (P, P)

H: hip, K: knee, W: walking, ST: walking stairs, S: sitting and standing up. N: NSAID, P = paracetamol. [a]Negative scores favour NSAIDs. [b]Scores range between 0 = no complaints and 6 = unbearable complaints for the total diary score and the individual main complaint, between 0 = no pain and 10 = very severe pain for the intensity of pain, and between 0 = very satisfied and 6 = very dissatisfied for the satisfaction with medication. [c]Confidence of general practitioner in medication of choice at the start of the study minus at the end of the study. Positive scores indicate that the general practitioner is more certain after the study. [d]Walking upstairs; walking downstairs. [e]Data of only four pairs could be compared.

perceived side effects reported increasing stomach complaints during treatment with both diclofenac and paracetamol (patient No 11). There were no differences in co-interventions between the NSAID and the paracetamol treatment periods. Patient No 10 had back pain and used more additional paracetamol as the study progressed. She finally withdrew from the study after eight weeks.

Treatment policy

Table 2 presents the treatment that the general practitioners thought to be best for each patient before the start and at the end of the study for the seven patients who completed the study. In general, after completion of the N-of-1 trials, the general practitioners were at least as certain about their treatment preferences as before the N-of-1 trials.

Follow-up

Three months after the end of the study, four of the six patients who had been recommended paracetamol treatment were taking NSAIDs for OA; one because of a misunderstanding about the recommended dosage of paracetamol (patient No 1), one because she preferred taking one small tablet of diclofenac rather than two large tablets of paracetamol (patient No 4), and two because of perceived lack of efficacy (patient Nos 7 and 9). Patient No 9 had been taking paracetamol for almost three months, when the complaints relating to OA deteriorated. This patient was therefore prescribed a high dose of NSAIDs at the time of the three-month evaluation, shortly after which the patient was prescribed both NSAIDs and paracetamol. Furthermore, one patient was taking both paracetamol and NSAIDs (patient No 3), because she thought it was a waste to throw away the NSAIDs she still had in her possession.

Discussion

In this study for each individual patient of the 13 selected, it was investigated whether paracetamol was as effective as NSAIDs in the treatment of pain and disability related to OA of the hip or knee. Six patients did not complete the study because of a perceived lack of efficacy, perceived side effects, concomitant disease, or loss of motivation. Seven

patients did complete the study. The results of the N-of-1 trials varied across these patients. Six patients were recommended to change to paracetamol treatment. All other patients (including the six who did not complete their trial) continued with NSAID treatment.

At the end of the follow-up period one of the six patients who were recommended to change to paracetamol treatment was using paracetamol and all others were using NSAIDs or both NSAIDs and paracetamol. Apparently, the use of NSAIDs was preferred by these patients, despite the equal effects of paracetamol. In a large survey [22] and in a few randomised trials [6,16] preferences of patients for either NSAIDs or paracetamol have been studied. About 60% of the patients in these studies preferred NSAIDs rather than paracetamol. However, this also implies that a considerable proportion (about 40%) indicated that paracetamol was better or as effective as NSAIDs. Preferring one drug to another is the result of several factors, including subjective benefit, side effects (actual or potential), doctor or patient beliefs and interactions, severity of disease, and ease of administration (for example, in our study one patient preferred taking one small tablet of diclofenac rather than two large ones of paracetamol).[22,23] To avoid wasting time and effort, these issues should be discussed and patients should be well informed about the objectives and consequences of the trials before starting an N-of-1 trial.

Considering the risk of side effects, the dosages were kept within the recommendations for chronic use of NSAIDs for OA published by the Dutch National Formulary.[9] Before the start of the study, however, higher dosages had sometimes been prescribed by the general practitioners. Three of the four patients who did not complete the study because of perceived lack of efficacy had been taking, before the start of the study, dosages of NSAIDs that were higher than the maximum dosages allowed in the study; this was not the case with any of the patients who completed the study. One other patient (No 13) who withdrew because of perceived lack of efficacy had used additional paracetamol/codeine before the study. These findings suggest that the withdrawal of these patients was due to subtherapeutic dosages of drugs during the study.

A series of N-of-1 trials will be able to demonstrate the potential variation in outcome across individual patients, but only if a large proportion of the participants complete their trial. However, dropout from an N-of-1 trial is an important outcome itself. As stated above, four of the six patients who did not complete their trial almost certainly

withdrew because of subtherapeutic dosages of drugs during the study. Therefore, we suggest that future research should consider comparing paracetamol with NSAIDs in the actual dosage taken by the patient just before the start of the study. In this way, we believe that dropout can be prevented in many cases.

March, *et al* and Nikles, *et al* conducted N-of-1 trials in patients with OA, using three pairs of treatment periods consisting of paracetamol and diclofenac [17] or paracetamol and ibuprofen [18]. The proportion of patients who did not complete their trial in our series (46%) is comparable with the results for these two other series (40% [17] and 43% [18]). Although a direct comparison is difficult (because of differences in selection criteria, drugs, and outcome measures), all three studies show that for some patients paracetamol is as effective as NSAIDs, but that there is a large variation in outcome among individual patients. The series by March, *et al* received some criticism of statistical analysis and the interpretation of the results.[24,25] March, *et al* considered paracetamol as the preferred drug if diclofenac was not significantly better than paracetamol. However, each patient received only three pairs of treatments, which limited the statistical power to detect treatment differences. Additionally, failure to find a significant difference in favour of diclofenac may not be regarded as proof of equivalence.

In our study each patient received five pairs of treatments. In contrast with March, *et al*, we used the individualised questionnaire developed by Guyatt, *et al* for N-of-1 trials.[19] This questionnaire examines the severity of symptoms as identified by the individual patient and scored by the patient on seven point ordinal scales. Our data were not suitable for quantitative analysis, partly owing to the non-normal distribution of the data. More importantly, we did not investigate whether one treatment was better than the other, but whether one treatment was as effective as the other. In other words, the N-of-1 trials were equivalence trials, rather than superiority trials. This implies that conventional significance testing could not be used for analysis and for formulating recommendations.[26] Therefore, we applied decision rules, which were formulated in advance. Although these rules can be considered to be arbitrary, for five of the seven patients who completed the study the differences between NSAIDs and paracetamol were very small and the recommendations were straightforward (Table 2). In two patients (Nos 6 and 7), the decision rules were very helpful in deciding on recommendations.

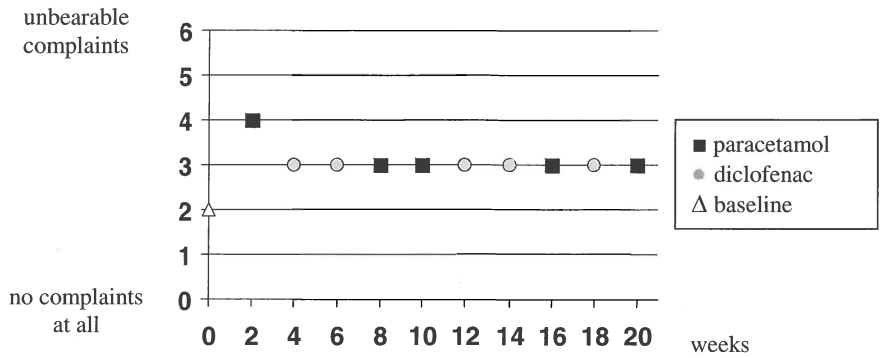

Figure 1a: patient number 9

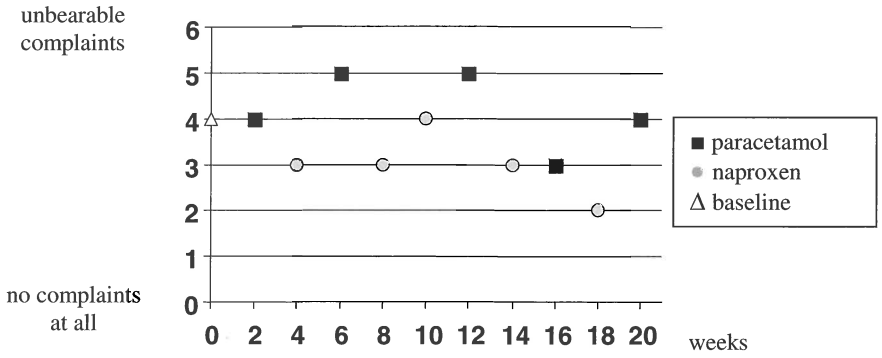

Figure 1b: patient number 6

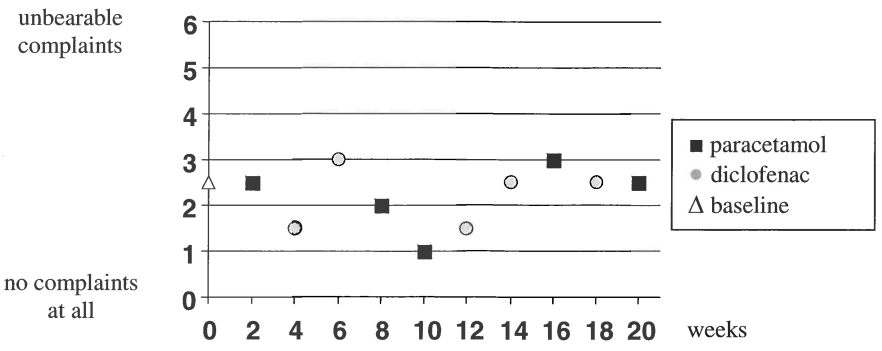

Figure 1c: patient number 4

Figure 1. Three typical plots for the total diary scores. No difference between NSAID and paracetamol (1a); NSAID better than paracetamol (1b); Wide range in scores - no association between outcome and type of drug (1c)

In conclusion, the results of the N-of-1 trials varied across patients (Figure 1). These findings support the assumption that the best treatment for OA differs in individual patients. However, there is, as yet, insufficient evidence about which factors predict which patient can or cannot successfully switch from NSAIDs to paracetamol. Therefore, an N-of-1 trial can be used to investigate which treatment is best for any specific individual patient. Practical reasons, such as the size or the number of tablets that needs to be taken, may play an important part in a patient's decision to switch from NSAIDs to paracetamol or not. Therefore, before the start of the N-of-1 trial, patients should be well informed about the objectives and consequences of the trial. When the results of the trial become available patients should be well instructed to follow the recommendations. In this way, unnecessary prolonged treatment with NSAIDs may be avoided in patients with complaints due to OA of the hip or knee.

Acknowledgements

We thank all the general practitioners and their patients who participated in the study. This study was supported unconditionally by Leo Pharma, The Netherlands.

References

1. Van Saase JL, van Romunde LK, Cats A, Vandenbroucke JP, Valkenburg HA. Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations. Ann Rheum Dis 1989;48:271-80.
2. Bradley JD, Brandt KD, Katz BP, Kalasinski LA, Ryan SI. Comparison of an antiinflammatory dose of ibuprofen, an analgesic dose of ibuprofen, and acetaminophen in the treatment of patients with osteoarthritis of the knee. N Engl J Med 1991;325:87-91.
3. Bradley JD, Brandt KD, Katz BP, Kalasinski LA, Ryan SI. Treatment of knee osteoarthritis: relationship of clinical features of joint inflammation to the response to a nonsteroidal antiinflammatory drug or pure analgesic. J Rheumatol 1992;19:1950-4.
4. Williams HJ, Ward JR., Egger MJ, Neuner R, Brooks RH, Clegg DO, Field EH, Skosey JL, Alarcon GS, Willkens RF et al. Comparison of naproxen and acetaminophen in a two-year study of treatment of osteoarthritis of the knee. Arthritis Rheum 1993;36:1196-1206.
5. Towheed TE, Hochberg MC. A systematic review of randomized controlled trials of pharmacological therapy in osteoarthritis of the knee, with an emphasis on trial methodology. Semin Arthritis Rheum 1997;26:755-70.
6. Pincus T, Koch CG, Sokka T, Lefkowitz J, Wolfe F, Jordan JM et al. A randomized double-blind, crossover clinical trial of diclofenac plus misoprostol versus acetaminophen in patients with osteoarthritis of the hip or knee. Arthr Rheum 2001;44:1587-98.
7. Anonymous. Do certain NSAIDs give less cause for gastrointestinal side effects than others? [In Dutch: Geven bepaalde NSAID' s minder aanleiding tot gastro intestinale bijwerkingen dan andere?] GeBu 1994;28:85-6.
8. MacDonald TM, Morant SV, Robinson GC, Shield MJ, McGilchrist MM, Murray FE, McDevitt DG. Association of upper gastrointestinal toxicity of non-steroidal anti-inflammatory drugs with continued exposure: cohort study. BMJ 1997;315:1333-7.
9. Van der Kuy A et al. Analgesics, antirheumatic agents and antigout agents. [In Dutch: Analgetica, antirheumatica en jichtmiddelen.] In: van der Kuy A ed. Farmacotherapeutisch Kompas 1999. Amstelveen: ZiekenfondsRaad, 1999:914-69.
10. Bijl D, Dirven-Meijer PC, Opstelten W, Raaijmakers AJ, Scholten RJPM, Eizinga WH, Goudswaard AN. National practice guidelines for non-traumatic kneeproblems in adults (Dutch College of Family Physicians). [In Dutch: NHG-Standaard Niet-Traumatische Knieproblemen bij Volwassenen.] Huisarts Wet 1998;41:344-50.
11. Pendleton A, Arden N, Dougados M, Doherty M, Bannwarth B, Bijlsma JW et al. EULAR recommendations for the management of knee osteoarthritis: report of a task force of the

Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis 2000;59:936-44.

12. Eccles M, Freemantle N, Mason J. North of England evidence based guideline development project: summary guideline for non-steroidal anti-inflammatory drugs versus basic analgesia in treating the pain of degenerative arthritis. The North of England Non-Steroidal Anti-Inflammatory Drug Guideline Development Group. BMJ 1998;317:526-30.
13. Holbrook AM, for Ontario Musculoskeletal Therapy Review Panel. Medical Treatment Guidelines for the Treatment of Osteoarthritis, Rheumatoid Arthritis, and Acute Musculoskeletal Injury. Toronto: Ontario Ministry of Health and Long-Term Care, 2000.
14. Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. American College of Rheumatology Subcommittee on Osteoarthritis Guidelines. Arthritis Rheum 2000;43:1905-15.
15. Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D, Keller J. A clinician' s guide for conducting randomized trials in individual patients. CMAJ 1988;139:497-503.
16. Solomon L, Abrams G. Orudis in the management of osteoarthritis of the knee. A double-blind trial. S Afr Med J 1974;48:1526-9.
17. March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. N of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. BMJ 1994;309:1041-5; discussion 1045-6.
18. Nikles CJ, Gasziou PP, Del Mar CB, Duggan CM, Clavarino A, Yelland MJ. Preliminary experiences with a single-patient trials service in general practice. Med J Aust 2000;173:100-3.
19. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The n-of-1 randomized controlled trial: clinical usefulness, our three-year experience. Ann Intern Med 1990;112:293-299.
20. Beurskens AJ, de Vet HC, Koke AJ, Lindeman E, van der Heijden GJ, Regtop W, Knipschild PG. A patient-specific approach for measuring functional status in low back pain. J Manipulative Physiol Ther 1999;22:144-8.
21. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, Scherr PA, Wallace RB. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994;49:M85-94.
22. Wolfe F, Zhao S, Lane N. Preference for nonsteroidal antiinflammatory drugs over acetaminophen by rheumatic disease patients. Arthrit Rheum 2000;43:378-85.

23. Pincus T, Swaeringen C, Cummings P, Callahan LF. Preference for non-steroidal antiinflammatory drugs versus acetaminophen and concomitant use of both types of drugs in patients with osteoarthritis. J Rheumatol 2000;27:1020-7.
24. Hawkey CJ. N of 1 trials in osteoarthritis, study was conceptually and statistically biased. [letter] BMJ 1995;310:666-7.
25. Senn S, Bakshi R, Ezzet N. Caution in interpretation needed. [letter] BMJ 1995;310:667.
26. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996;313:36-9.

Chapter 4

Efficacy of temazepam in frequent users: a series of N-of-1 trials

In press as:

Wegman ACM, Van der Windt DAWM, Bongers M, Twisk JWR, Stalman WAB, De Vries ThPGM. Efficacy of temazepam in frequent users: a series of N-of-1 trials. Fam Pract.

Abstract

Background: Benzodiazepines are frequently prescribed for sleep disturbances. However, benzodiazepines are associated with side effects, and may be ineffective when used for a prolonged period of time.

Objectives: To investigate for individual patients whether placebo is as effective as temazepam, or whether 10 mg was as effective as 20 mg temazepam, and whether these results influenced their future temazepam use.

Methods: A series of randomised double-blind N-of-1 trials were conducted in general practices in The Netherlands for patients who were using temazepam regularly. Each patient received five pairs of treatments consisting of one week of temazepam (10 or 20 mg) and one week of the control intervention (placebo or 10 mg temazepam). Per pair, the sequence of treatments was randomised. Main outcome measures were: time to fall asleep, and the individual main complaint.

Results: Twelve out of 15 patients completed their trial. In three patients there was no difference, in five a large difference, and in four a small difference in favour of temazepam. At follow-up, seven patients had stopped or reduced their temazepam use.

Conclusions: The results regarding the efficacy of temazepam varied across patients. N-of-1 trials seem to be valuable in patients who are motivated to stop or reduce their temazepam use. They clearly demonstrate the efficacy of temazepam, and may give patients additional confidence to discontinue regular hypnotic use. The value of N-of-1 trials for patients who are less motivated is unclear, as the size of treatment effect does not seem to influence future hypnotic use.

Introduction

Benzodiazepines are frequently prescribed for sleep disturbances.[1] In 1998, the annual prevalence of benzodiazepine use was estimated at 12.2% in The Netherlands. At least one third of all users were long-term users.[2] However, there is considerable doubt about whether benzodiazepines are still effective when used for a prolonged period of time.[3-5] Additionally, use of benzodiazepines is associated with adverse cognitive effects and an increased risk of motor vehicle accidents, falls and fractures, especially among the elderly.[6-8] Furthermore, benzodiazepines may interact with other medication, such as antidepressants and antiepileptic drugs, resulting in an intensified hypnotic effect.[9] Therefore, in practice guidelines for the management of insomnia it is recommended to prevent or stop long-term use of benzodiazepines.[3,5]

However, attempts made by general practitioners (GPs) and their patients to stop long-term use of benzodiazepines are often unsuccessful. Success rates varying from 13% to 59% have been reported.[10-15] The success of such an attempt may be influenced by multiple factors, such as psychological and social factors, dependence on the hypnotic, the coaching given by the GP, and variation in the effectiveness of the hypnotic across individuals. But probably the most important obstacle for successful stopping is the awareness of the patient of the abstinence of his or her daily dose of benzodiazepines. N-of-1 trials may provide more insight into the question whether the use of hypnotics can be successfully reduced in individual patients without loss of quality and quantity of sleep. Because in N-of-1 trials both the patient and the GP are blinded for the sequence of treatments, the potential influence of expectations with regard to the effectiveness of the hypnotic is precluded.

N-of-1 trials are carried out with one patient. The patient is his or her own control, and receives alternately the intervention treatment and the control treatment. The sequence of treatments is randomised.[16] N-of-1 trials may help a GP to decide on a treatment policy when there is doubt regarding the effectiveness of medication for an individual patient. N-of-1 trials are especially useful in patients with chronic conditions, and for interventions that have a rapid onset and stop acting soon after discontinuation.[16] Several N-of-1 trials have been carried out and their feasibility in clinical practice seems to be promising.[17,18,19] Since the purpose of N-of-1 trials is to evaluate the effects of

treatments for each individual patient (and not to estimate the average effect for a larger population), the size of a series of N-of-1 trials is of minor importance. Nevertheless, a series of N-of-1 trials will be able to demonstrate the potential variation in outcome across individuals.

The first objective of this study was to investigate for individual patients in general practice whether placebo is as effective as temazepam (10 mg or 20 mg), or, in some patients, whether 10 mg temazepam is as effective as 20 mg temazepam. The second objective was to investigate whether presenting the personal results of the N-of-1 trial to each individual patient influenced their future use of temazepam.

Methods

Patient selection

Nine GPs from six different towns in The Netherlands participated in the study. Between April and November 2001, they were asked to select patients from their medical records who met the following criteria: regular use of 10 mg or 20 mg temazepam as a hypnotic for at least 4 nights a week during the past 2 months; at least 18 years of age; and no contra-indications for benzodiazepines. Additionally, the patients had to be able to fill in a questionnaire in Dutch and they had to be able to visit the practice.

Design

The selected patients received information about the study, and written informed consent was obtained. The study comprised of a series of N-of-1 trials with a duration of ten weeks. Before the start of their trial, patients were asked about their motivation to participate. To standardise the knowledge on sleep hygiene for all patients, written recommendations for sleep hygiene [3] were given to and discussed with the patient by the GP before the patients received any trial medication. During the study, each patient received five pairs of treatments. Each pair consisted of a one-week period of treatment A and a one-week period of treatment B (Figure 1). A patient taking 10 mg temazepam before the start of the study received placebo (treatment A) versus 10 mg temazepam (treatment B) during the study; a patient taking 20 mg temazepam before the start of the

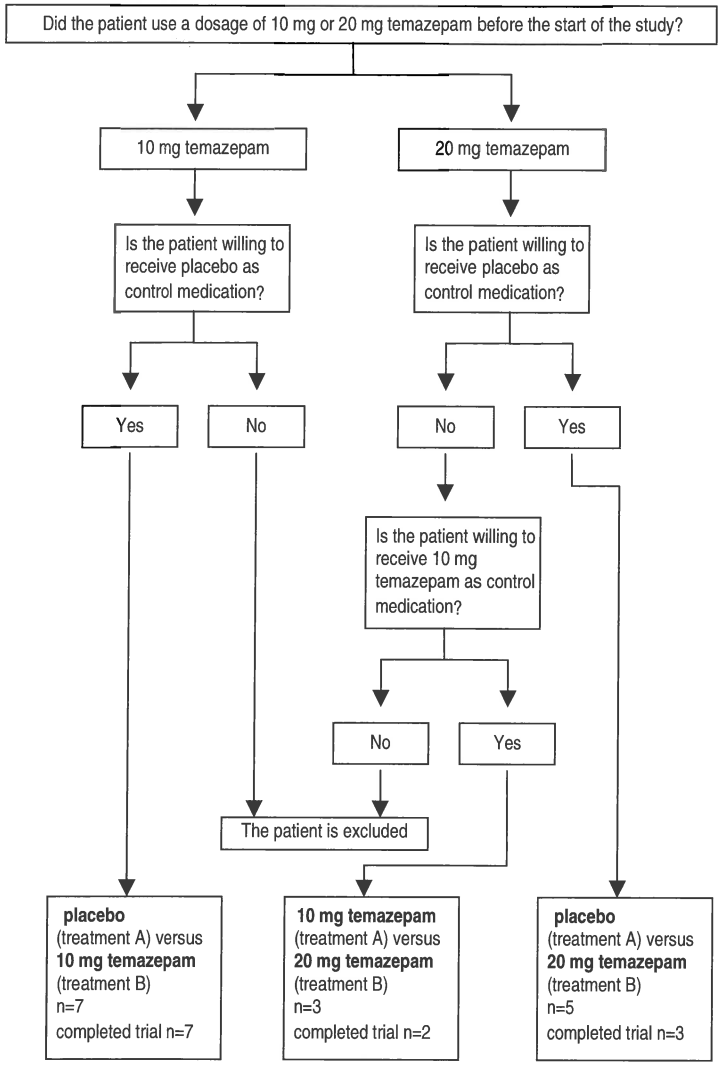

Figure 1. Flow diagram to determine the trial medication per individual

study received either placebo or 10 mg temazepam (treatment A) versus 20 mg temazepam (treatment B) during the study, depending on the patient' s willingness to either reduce the dosage or to stop completely. Per pair of treatments, the sequence of the two treatments was randomised.

The random sequence was prepared in advance by the hospital pharmacist for each patient separately. The patient, the GP and the investigator were blinded for the sequence of treatments. To ensure blinding, identical tablets with regard to colour, smell and taste were used as placebo. If 10 mg temazepam was compared to 20 mg temazepam, additional placebo tablets were given in the 10 mg temazepam treatment periods, so that the patient received 2 tablets each night in all treatment periods. Patients were not instructed to take the medication daily, but every night they decided for themselves whether or not to take the tablets. The patients received the tablets in special boxes, in which each section was coded with a date and contained the tablets for that date. The patients were instructed to leave the tablets they did not take in the boxes. In addition, the patients were asked to indicate each day in a diary whether or not they had taken any hypnotics. The study was approved by the Ethics Review Board of the VU University Medical Center.

Outcome assessment

Outcomes were assessed by means of daily diaries. The primary outcome measures were:

1) the individual main outcome for which the patient could select among the following outcomes: time to fall asleep (minutes); the patient' s perception of the night' s sleep (total sleep time sufficient or not); the total sleep time (hours); the number of times awake during the night; the duration of the longest period awake during the night (minutes); and the number of complaints in the daytime (selected from a set of 11, which could indicate withdrawal symptoms or side effects); and
2) independent of the individual main outcome, the time to fall asleep (minutes).

All other outcome measures mentioned above were considered as secondary outcome measures for each individual patient. Furthermore, information was collected on co-

medication, alcohol intake and afternoon naps. After analysis, plots of their scores on primary and secondary outcome measures were shown to the patients and discussed with them. Three months later the patients were contacted by telephone and asked how many times, and in which dosage, they had taken temazepam during the previous 2 weeks.

Analysis

Since N-of-1 trials were conducted, all outcome measures were analysed for each patient separately. Because of the non-normal distribution of the data, median scores were calculated for all continuous outcomes for each treatment period (Figure 2). Furthermore, for each treatment period percentages were calculated for the number of nights with a sufficient total sleep time and for the number of days with at least one complaint in the daytime. So- called 'bad' nights were defined as nights in which the time to fall asleep was > 60 minutes, or in which the total sleep time was ≤ 5 hours. The percentage of 'bad' nights was calculated for each treatment period. Finally, differences in the scores between the compared treatments were calculated (Figure 2).

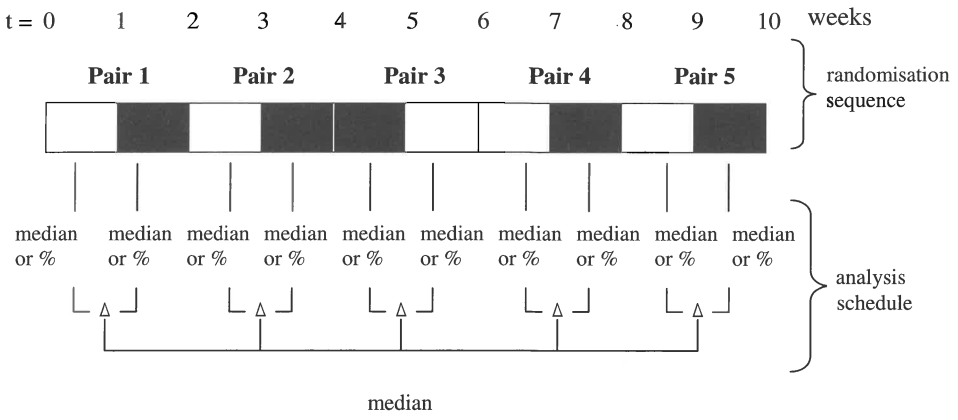

Figure 2. Example of randomisation sequence for one patient comparing placebo (treatment A) and 10 mg temazepam (treatment B), and analysis schedule

A priori, cut-off points were defined for the primary outcome measures. A median difference in the time to fall asleep, the total sleep time, or the longest period awake during the night, of at least 30 minutes in favour of temazepam was considered to be a large effect; 5 to 30 minutes a small effect; and < 5 minutes no effect of temazepam. For differences in the number of times awake during the night and the number of complaints in the daytime the cut-off points for large, small and no effect of temazepam were >1, between 0 and 1, and 0, respectively. Finally, the cut-off points for the number of nights with sufficient sleep were a difference of at least 29% (2 nights in a week), between 0% and 29%, and 0%, respectively for large, small, and no effect of temazepam.

All analyses were carried out according to the intention-to-treat principle. For patients who took less than 75% of the trial medication, an alternative analysis was carried out, including only those days on which the trial medication or additional benzodiazepines had been taken (on-treatment analysis).

Assessment of the plots of the primary and secondary outcome measures showed that there were no carry-over effects, i.e. there were no consistently better results in the first nights (compared to the latter nights) of a control intervention period following a period of (highest dose of) temazepam use or vice versa. Therefore, the results of all 7 days of each treatment period were entered in the analysis.

Results

Subjects

Fifteen patients were selected for participation: ten women and five men (Table 1). Their age ranged between 45 and 78 years (median 62 years). Before the start of the study, seven patients were taking 10 mg temazepam and eight were taking 20 mg temazepam. Of the eight patients who were taking 20 mg temazepam, five received placebo versus 20 mg temazepam, and three received 10 mg temazepam versus 20 mg temazepam during the study.

Table 1. Patient characteristics of 15 participants in N-of-1 trials on the efficacy of temazepam

Patient No	Sex	Age	Use of temazepam before start of trial	Total duration of hypnotic use (any type or dose)	Trial medication (mg temazepam)
1	M	47	7 nights / week 20 mg for 5 months	5 months	10 versus 20
2	F	78	7 nights / week 10 mg for 8.5 years	8.5 years	0 versus 10
3	F	48	7 nights / week 20 mg for 11 years	11 years	0 versus 20
4	M	62	6 to 7 nights / week 20 mg for 2 years	6.5 years	0 versus 20
5	M	74	4 to 5 nights / week 10 mg for 8 years	12 years	0 versus 10
6	F	56	7 nights / week 20 mg for 6 years	7 years	0 versus 20
7	F	57	5 nights / week 10 mg for 15 years	15 years	0 versus 10
8	F	45	7 nights / week 10 mg for 8 years	8 years	0 versus 10
9	F	72	7 nights / week 10 mg for 9 years	30 years with intervals	0 versus 10
10	F	73	7 nights / week 20 mg for years	For years	10 versus 20
11	M	70	7 nights / week 10 mg for 6 months	6 months	0 versus 10
12	F	71	4 nights / week 10 mg for 8 years	12 years	0 versus 10
13	F	56	7 nights / week 20 mg for 3 years	3.5 years	10 versus 20
14	M	76	6 nights / week 20 mg for 4 months	4 months	0 versus 20
15	F	53	7 nights / week 20 mg for 2 years	2 years	0 versus 20

M: male, F: female

Three patients (patient Nos 13, 14, and 15) did not complete the study: one because of nausea after a period of 20 mg temazepam (patient No 14 after a few days), and two because of insomnia (No 13 after a few days, after a period of the lowest dose of temazepam; No 15 after 5 weeks, after a period of 20 mg temazepam). Patient No15 took additional benzodiazepines during the last 12 days before she stopped the trial. All three patients took 20 mg temazepam before the start of the study.

Efficacy of temazepam

Twelve patients completed their 10-week trial period. Before the patients were informed about the results of their N-of-1 trial, they were asked whether they had any idea when they had received the (highest dose of) temazepam. Blinding turned out to be successful: none of the twelve patients had noticed that they had taken temazepam as well as the control treatment for periods of one week.

For all patients, there were no or minor differences for co-medication, alcohol intake and the number of afternoon naps between temazepam and the control intervention. The results of the N-of-1 trials regarding the efficacy of temazepam varied across the patients (Table 2). For three patients (Nos 1, 2, and 3), the primary outcomes (minutes to fall asleep and the individual main outcome) showed no difference between the two treatments compared. For three other patients (Nos 10, 11, and 12), the time to fall asleep was at least 30 minutes in favour of (the highest dose of) temazepam, and the individual main outcome favoured (the highest dose of) temazepam. For five other patients (Nos 4, 5, 6, 7, and 8), the primary outcomes favoured temazepam to a lesser extent. Finally, in one patient (patient No 9), the time to fall asleep was only 5 minutes in favour of temazepam, whereas the duration of the longest period awake during the night (the individual main outcome for this patient) was 40 minutes in favour of temazepam. For all patients, there were no or only small differences for the number of complaints in the daytime between temazepam and the control intervention, which indicates no differences with regard to possible withdrawal symptoms or side effects.

For five patients (Nos 5, 6, 7, 10, and 12) the percentage of 'bad' nights was at least 29% (2 nights/week) in favour of temazepam for one or both outcomes. For all other patients the percentage of 'bad' nights showed no or only minor differences between the two treatments (< 2 night/week).

On-treatment analysis

Ten of the twelve patients who completed the study took at least 94% of the trial medication. The other two patients took less than 75% of the medication (patient No 2, 60%; patient No 7, 43%), so on-treatment analysis was carried out. For patient No 2 the on-treatment analysis showed similar results as the intention-to-treat analysis for the individual main outcome, but the time to fall asleep was 15 minutes in favour of placebo

Table 2. Results for the 12 patients who completed their N-of-1 trial. (negative scores favour (highest dose of) temazepam)

Patient No	Trial medication (mg temazepam)	Median difference between treatments (range)				
		Minutes to fall asleep	% nights with sufficient sleep	Minutes slept	No. of times awake during the night	Time awake during the night (minutes)
1	10 versus 20	**0** * (- 17.5; 35)	0 (- 29; 29)	0 (- 30; 60)	0 (- 1; 1)	0.25 (- 12.5; 7.5)
2	0 versus 10	0 (0; 0)	**0** * (- 14; 0)	- 15 (- 30; 30)	0 (0; 0)	0 (- 5; 0)
3	0 versus 20	**0** * **(-20; 0)**	0 (- 40; 17)	- 30 (- 90; 60)	0 (0; 0)	0 (0; 0)
4	0 versus 20	- 15 (- 15; 15)	**2** * (- 14; 29)	0 (0; 60)	0 (- 1; 0)	- 7.5 (- 15; 0)
5	0 versus 10	- 12,5 (- 125; 0)	- 36 (- 43; - 16)	- 45 (- 90; - 30)	0.5 (- 1; 1)	**- 2.5** * (- 32.5; 0)
6	0 versus 20	- 5 (- 45; 10)	- 21 (- 57; - 7)	- 75 (- 120; - 30)	_ †	**- 15** * (- 30; - 5)
7	0 versus 10	- 15 (- 30; 30)	**- 14** * (- 29; 36)	- 60 (- 60; 15)	- 0.25 (- 1; 0)	- 5 (- 45; 5)
8	0 versus 10	- 5 (- 30; 0)	**- 26** * (- 43; 14)	- 30 (- 60; 45)	0 (- 1; 0.5)	- 5 (- 7.5; 7.5)
9	0 versus 10	- 5 (- 12.5; 5)	- 43 (- 43; - 12)	- 60 (- 180; 0)	- 1 (- 1.5; 0)	**- 40** * (- 205; 20)
10	10 versus 20	**- 30** * (- 45; 10)	9.6 (- 14; 29)	0 (- 45; 60)	0 (- 0.5; 1)	- 2.5 (- 20; 80)
11	0 versus 10	- 30 (- 60; 0)	- 43 (- 46; - 12)	- 45 (- 90; 15)	**- 1** * (-1; 0.5)	0 (- 17.5; 15)
12	0 versus 10	**- 45** * (- 90; 10)	- 43 (-55; 19)	- 90 (- 120; 30)	- 0.5 (- 1; 1)	- 5 (- 10; - 2.5)

* The results of the individual main outcome are presented in bold; †Insufficient information

according to the on-treatment analysis. For patient No 7 the results of both primary outcomes were more strongly in favour of temazepam according to the on-treatment analysis. The time to fall asleep was 49 minutes shorter during temazepam use and the percentage of nights with sufficient total sleep time was 30% in favour of temazepam, according to the on-treatment analysis.

Efficacy of temazepam and follow-up

After presenting the results of the primary and secondary outcome measures to the patients, they were asked about their intentions with regard to future temazepam use (Table 3). All three patients for whom the primary outcomes showed no difference between the treatments (Nos 1, 2, and 3) intended to stop or reduce the use of temazepam. At three months follow-up one patient (no. 1) again used temazepam as before the study. Three of the four patients who showed a small benefit of temazepam (Nos 4, 6, and 8) intended to stop or reduce the use of temazepam, but only two persisted in this intent at three months follow-up. Finally, three of the five patients with large effects of temazepam (Nos 9, 10, and 11) had stopped or reduced the use of temazepam at three months follow-up as intended shortly after the trial. In sum, seven of the 12 patients had stopped (3 patients) or reduced (4 patients) temazepam use at follow-up. There seemed to be only a weak association between the efficacy of temazepam and the intentions regarding future temazepam use, and no association between the efficacy of temazepam and the actual hypnotic use at follow-up.

Motivation to participate

Nine of all 15 patients indicated at baseline participated in the study because they wanted to stop or reduce their use of temazepam. Five patients participated for the benefit of science and one because "he was asked to participate by the GP's assistant" (Table 3). This last patient (no. 14) did not complete his trial. Of the nine patients who were motivated to reduce their temazepam use, two did not complete their trial (Nos 13 and 15), six stopped or reduced the use of temazepam at three months follow-up (Nos 2, 3, 6, 8, 9, and 11), and only one used temazepam as before the study (No 7). Among the five patients who participated to benefit science, one patient had reduced the use of

Table 3. Results for the 15 participants in N-of-1 trials: motivation to participate, efficacy of temazepam, intended future use of hypnotics, and hypnotic use three months after presenting the outcomes

Patient No	Motivation to participate in N-of-1 trial [a]	Efficacy of temazepam	Intended medication use after trial [b]	Medication after 3 months [b]
1	S	No effect	Reduce dose and freq	As before study
2	P	No effect	Reduce freq	Reduced freq
3	P	No effect	Stop	Stopped
4	S	Small effect	Reduce freq	As before study
5	S	Small effect	As before study	As before study
6	P	Small effect	Reduce freq	Reduced freq
7	P	Large effect (on-treatment analysis)	As before study	As before study
8	P	Small effect	Reduce freq or stop	Reduced freq
9	P	Large effect	Stop	Stopped
10	S	Large effect	Reduce dose and freq	Reduced dose and freq
11	P	Large effect	Stop	Stopped
12	S	Large effect	As before study	As before study
13 [c]	P	-	-	-
14 [c]	O	-	-	-
15 [c]	P	-	-	-

[a] P=Patient motivated to stop or reduce temazepam use, S=for the benefit of science, O=other (asked by the GP's as sistant); [b] freq=frequency;
[c] Patient Nos 13, 14, and 15 did not complete their trial

temazepam (patient No 10) and four used temazepam as before the study (Nos 1, 4, 5, and 12) at the three months follow-up.

Discussion

In a series of randomised, blinded N-of-1 trials we investigated the efficacy of temazepam in frequent users of hypnotics. Twelve of the fifteen patients (80%) completed the 10-week trial period. The results among the 12 completers varied from no difference to large differences in favour of (the highest dose of) temazepam. The results did not seem to be associated with the dosage of temazepam used during the trial. For example, among the three patients who experienced no effect of temazepam, one patient received placebo versus 10 mg temazepam, another 10 mg versus 20 mg temazepam, and the third placebo versus 20 mg temazepam.

Large-scale trials have shown that attempts to stop temazepam use are only successful in 13 to 59% of patients.[10-15] There is little evidence regarding the factors that may predict which patients can or cannot successfully reduce the use of temazepam. Age under 65 [10] has been reported to be associated with a successful attempt, but this evidence is not consistent across studies. In our series age did not seem to influence outcomes. Cormack, *et al* demonstrated that a relatively low consumption of benzodiazepines [11] was associated with a successful attempt. However, in our series, none of the three patients taking temazepam five or less times a week at baseline, intended to stop or reduce temazepam use after the trial. Due to the less frequent use of temazepam compared to the other patients in the study, it may be harder to achieve a reduction in the frequency of temazepam intake.

N-of-1 trials demonstrate the efficacy of temazepam in each individual patient, and may thereby provide a good indication of the possibility to reduce hypnotic use. However, in our series, the actual results of the trials did not seem to be a strong predictor of future use of temazepam. Patients decided either to stop, reduce, or continue the use of hypnotics regardless of the effect of temazepam shown by their trial. While the influence of treatment effect appeared to be limited, the motivation of the patients to participate in the study seemed to be of greater importance. Six of the seven patients who completed

their trial and had indicated that they participated in the study because they wanted to reduce temazepam intake, successfully reduced temazepam use during follow-up. In contrast, out of the five completers who participated for the benefit of science only one reduced the intake of hypnotics.

The potential influence of motivation to change behaviour has been demonstrated in other research. In an intervention study among 72 long-term benzodiazepine users Morrison [12] found that it was significantly more likely that the patient would stop taking the drug if they originally wished to do so.

The question arises whether the patients with a (strong) internal motivation needed the N-of-1 trial to change hypnotic intake. In our opinion, the participation in an N-of-1 trial may strengthen the intent to change hypnotic use. Eight of the nine patients who intended to reduce temazepam intake stated that the N-of-1 trial had contributed to this decision. By participating in a trial the patients are given more attention to their use of temazepam and become more aware of their frequent hypnotic use. This attention itself may be of more importance than the results of the individual trials to stimulate the patients to stop or reduce the temazepam intake. In addition, for patients who are willing to reduce temazepam intake, but are somewhat afraid to do so, an N-of-1 trial may be more effective than an unblinded attempt to reduce intake. The awareness of abstinence of temazepam may cause insomnia, which may negatively influence the success of the attempt. The fact that the blinded design of an N-of-1 trial may facilitate the attempt to reduce drug use is illustrated by a statement made by patient No 3:

> *"For a long period of time (before the study), I wanted to stop, but I was uncertain whether I could hang on. Because of the project I had to persist. All the time during the trial, I slept reasonably well, and it made no difference what I took (temazepam or placebo). Therefore I stopped (taking sleep medication)."*

At follow-up, seven of the 12 patients who completed the trial, had stopped or reduced their temazepam intake. Therefore, double-blinded N-of-1 trials may be viewed as an instrument for GPs to reduce the unnecessary temazepam use of their patients, especially for those who are willing but somewhat afraid to do so.

Blinding of patients was successful, unexpectedly even for patients in whom temazepam favoured placebo (or lower dosage of temazepam) to a larger extent. This may be explained by the fact that quality and quantity of sleep varied across nights regardless of the type of treatment, meaning that patients were not always sleeping well when taking temazepam, and not always sleeping badly when taking the control intervention. The latter can also be concluded from the range in outcomes within the patients presented in Table 2. Furthermore, for all patients, there were no or only small differences for the number of complaints in the daytime (possible withdrawal symptoms or side effects) between temazepam and the control intervention, which helped to ensure blinding.

The dropout rate in our study was relatively low (3 of the 15 patients). To prevent dropout due to the length of the trial period, we wanted to keep the trial period as short as possible. However, this also implies that the power, which depends on the number of treatment pairs, is limited. These aspects should be weighed against each other when designing an N-of-1 trial.[20] The power of a study is also related to the type of significance testing. Although significance testing may be possible in N-of-1 trials, our data were not suitable for quantitative analysis, partly due to the non-normal distribution of the data. More importantly, we did not investigate whether one treatment was better than the other, but whether one treatment was equally effective as the other. In other words, the N-of-1 trials were equivalence trials, rather than superiority trials. This implies that conventional significance testing could not be used for analysis.[21] This topic was also addressed in a reaction to a series of N-of-1 trials by March, *et al*, in which it was pointed out that failure to find a significant difference in favour of a treatment may not be regarded as proof of equivalence.[22,23] As a consequence, we decided to formulate a priori cut-off points, which were very helpful to determine the magnitude of treatment effect, and to identify the variation in efficacy of temazepam across patients.

In conclusion, N-of-1 trials provide clear evidence about the impact of temazepam on the quality and quantity of sleep in individual frequent users of hypnotics. In patients who are motivated to stop or reduce their temazepam intake, participation in the trial may give additional confidence to successfully reduce temazepam use. The value of N-of-1 trials for patients who are less motivated is unclear, as the size of treatment effect does not

seem to influence future use of hypnotics. For less motivated patients experiencing small or no effect of temazepam, additional coaching by the GP, and education about the potential risks and disadvantages of long-term or frequent use of hypnotics may offer possibilities to reduce their unnecessary temazepam use. For patients experiencing large benefits and no potential disadvantages of temazepam, continued use of temazepam could be justified. Our study demonstrated that there is a realistic potential that N-of-1 trials are helpful in reducing benzodiazepine use in primary care. The next step could be to study, in a randomised clinical trial, whether N-of-1 trials are more successful than other interventions in primary care to reduce unnecessary temazepam use.[24] Finally, the design and execution of N-of-1 trials is time-consuming. Therefore, also the applicability of N-of-1 trials in general practice needs further study, in order to assess the possibilities of implementing these trials in daily practice.

Acknowledgements

We wish to thank all the general practitioners and their patients who participated in the study. This study was supported unconditionally by Leo Pharma, The Netherlands.

References

1. Kruse WH. Problems and pitfalls in the use of benzodiazepines in the elderly. Drug Saf 1990;5:328-44.
2. Gorgels WJMJ, Oude Voshaar RC, Mol AJJ, Breteler MHM, Lisdonk EH van de, Zitman FG. Long-term use of benzodiazepines. [In Dutch: Het langdurig gebruik van benzodiazepinen.] Ned Tijdschr Geneeskd 2001;145:1342-6.
3. Knuistingh Neven A, Graaff WJ de, Lucassen PLBJ et al. National practice guideline for insomnia and hypnotics (Dutch College of General Practitioners). [In Dutch: NHG-Standaard Slapeloosheid en slaapmiddelen.] In: Rutten GEHM, Thomas S (eds.). NHG-Standaarden voor de huisarts. Utrecht: Bunge 1993:264-77.
4. Lader MH. Limitations on the use of benzodiazepines in anxiety and insomnia: are they justified? European Neuropsychopharmacology 1999;9(Suppl 6):S399-S405.
5. Blow FC (consensus panel chair). Substance Abuse Among Older Adults. Treatment Improvement Protocol (TIP) Series 26. Rockville: U.S. Department of Health and Human Services (DHHS), 1998.
6. Thomas RE. Benzodiazepine use and motor vehicle accidents. Systematic review of reported association. Can Fam Physician 1998;44:799-808.
7. Ray WA, Fought RL, Decker MD. Psychoactive Drugs and the Risk of Injurious Motor Vehicle Crashes in Elderly Drivers. Am J Epidemiol 1992;136:873-83.
8. Herings RM, Stricker BH, Boer A de, Bakker A, Stuurmans F. Benzodiazepine and the risk of falling leading to femur fractures. Dosage more important than elimination half-life. Arch Intern Med 1995;155:1801-7.
9. Van Loenen AC, ed. Central nervous system (mental disorders). [In Dutch: centrale zenuwstelsel (psychische aandoeningen)] In: van Loenen AC, ed. Farmacotherapeutisch Kompas 2002. Amstelveen: College voor zorgverzekeringen (CVZ), 2001:53-135.
10. Holden JD, Hughes IM, Tree A. Benzodiazepine prescribing and withdrawal for 3234 patients in 15 general practices. Fam Pract 1994;11:358-362.
11. Cormack MA, Owens RG, Dewey ME. The effect of minimal interventions by general practitioners on long-term benzodiazepine use. J R Coll Gen Pract 1989;39:408-11.
12. Morrison JM. Audit and follow-up of chronic benzodiazepine tranquillizer use in general practice. Fam Pract 1990;7:253-7.
13. Morrice A, Iliffe S. Advising patients on their benzodiazepine use. Br J Gen Pract 1990;40:83.

14. Cormack MA, Sweeney KG, Hughes-Jones H, Foot GA. Evaluation of an easy, cost-effective strategy for cutting benzodiazepine use in general practice. Br J Gen Pract 1994;44:5-8.
15. Hopkins DR, Sethi KBS, Mucklow JC. Benzodiazepine withdrawal in general practice. J R Coll Gen Pract 1982;32:758-762.
16. Guyatt G, Sackett D, Adachi J et al. A clinician' s guide for conducting randomized trials in individual patients. CMAJ 1988;139:497-503.
17. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The n-of-1 randomized controlled trial: clinical usefulness, our three-year experience. Ann Intern Med 1990;112:293-299.
18. Larson EB, Ellsworth AJ, Oas J. Randomized clinical trials in single patients during a 2-year period. JAMA 1993;270:2708-12.
19. Patel A, Jeaschke R, Guyatt GH, Keller JL, Newhouse MT. Clinical usefulness of n-of-1 randomized controlled trials in patients with nonreversible chronic airflow limitation. Am Rev Respir Dis 1991;144:962-4.
20. Campbell MJ. Commentary: statistical aspects. BMJ 2004;328:506
21. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996;313:36-9.
22. March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. N of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. BMJ 1994;309:1041-5; discussion 1045-6.
23. Senn S, Bakshi R, Ezzet N. Caution in interpretation needed. [letter] BMJ 1995;310:667.
24. Mahon J, Laupacis A, Donner A, Wood T. Randomised study of n of 1 trials versus standard practice. BMJ 1996;312:1069-74.

Chapter 5

Conducting research in individual patients: lessons learnt from two series of N-of-1 trials

Submitted as:
Wegman ACM, Van der Windt DAWM, Stalman WAB, De Vries ThPGM. Conducting research in individual patients: lessons learnt from two series of N-of-1 trials.

Abstract

Background: Double-blind randomised N-of-1 trials (N-of-1 trials) may help with decisions concerning treatment when there is doubt regarding the effectiveness and suitability of medication for individual patients. The patient is his or her own control, and receives during several periods of time the experimental and the control treatment in random order. Reports of N-of-1 trials are still relatively scarce, and the research methodology is not as firmly established as that of RCTs. Recently, we have conducted two series of N-of-1 trials in general practice.

Objectives: To describe the difficulties we encountered during these series, and to discuss our solutions to these difficulties.

Methods: Before, during, and after data-collection, difficulties regarding outcome assessment, analysis of the results, the withdrawal of patients, and the follow-up had to be dealt with. These difficulties are described and our solutions are discussed.

Results: To prevent or anticipate difficulties in N-of-1 trials, it is important to individualise the outcome measures, and to carefully consider the objective, type of randomisation and the analysis. It is recommended to use the same dosages and dosage forms that the patient used before the trial, to start the trial with a run-in period, to formulate both general and individualised decision rules regarding the efficacy of treatment, to adjust treatment policies immediately after the trial, and to provide adequate instructions and support if the treatment is adjusted.

Conclusions: The design of an N-of-1 trial has to be tailored to the specific characteristics of each individual patient and the underlying medical problem. In this way, N-of-1 trials may be of great help when deciding on drug treatment for individual patients.

Introduction

In medical care, deciding on the optimal treatment for individual patients is one of the main concerns of the physician. Scientific evidence, recommendations from guidelines, medical expertise, patient preferences, and personal experiences all contribute to this decision process. Randomised controlled trials (RCTs) are generally considered to provide the strongest evidence for the efficacy of treatment. However, RCTs are designed to estimate an average treatment effect in a specific population.[1] In daily practice the physician has to determine the extent to which this average effect will apply to an individual patient. For instance, a patient who consults a physician may not be of a similar age, may have additional co-morbidity and medication, or may be interested in a different outcome, compared to the subjects studied in the related RCT.

N-of-1 trials are carried out with one patient only, and may help to decide on a treatment policy when there is doubt regarding the effectiveness and suitability of medication. In contrast to RCTs, N-of-1 trials do not assess what is best on average for a whole population, but what is best for an individual patient. The patient is his or her own control, and receives during several periods of time the experimental and the control treatment in random order. If possible, the patient, the physician and the researcher are blinded for the sequence of treatments.[2]

Reports of N-of-1 trials in general practice are still relatively scarce, and the research methodology is not as firmly established as that of RCTs. Recently, we have conducted two series of N-of-1 trials (Table 1), one in patients with osteoarthritis of the hip or knee (series A), and one in long-term users of temazepam (series B).[3,4] Before, during, and after data-collection, a number of difficulties had to be dealt with. The aim of this paper is to describe the difficulties we encountered during these series and to discuss our solutions to these difficulties. Successively, difficulties with regard to the outcome assessment, the analysis of the results, the withdrawal of patients, and the follow-up will be discussed.

Table 1. Characteristics of two series of N-of-1 trials conducted by our research group

	Series A [3]	Series B [4]
Setting	General practice	General practice
Subjects	Patients who had been taking NSAIDs regularly in the treatment of pain and disability related to osteoarthritis of the hip or knee.	Long-term users of temazepam (10 or 20 mg) for sleep disturbances.
Objective	1) Is paracetamol as effective as NSAIDs? 2) Is medication-use influenced by presenting the personal results to each individual patient?	1) Is placebo as effective as temazepam (10 or 20 mg), or, in some patients, is 10 mg temazepam as effective as 20 mg temazepam? 2) Is medication-use influenced by presenting the personal results to each individual patient?
Primary outcome measures	1) individual main complaints 2) intensity of pain	1) individual main complaint 2) time to fall asleep
Treatment pairs	2 weeks of paracetamol and 2 weeks of NSAIDs	1 week of placebo and 1 week of 10 mg temazepam, or 1 week of placebo and 1 week of 20 mg temazepam, or 1 week of 10 mg temazepam and 1 week of 20 mg temazapam
Total trial period	20 weeks (5 pairs of treatment periods)	10 weeks (5 pairs of treatment periods)
Sequence of treatments	Randomisation within each treatment pair	Randomisation within each treatment pair
Blinding	The patient, the investigator and the GP were blinded for the sequence of treatments.	The patient, the investigator and the GP were blinded for the sequence of treatments.

Outcome assessment

Problem: Which outcome measures should be used?

In a series of N-of-1 trials it seems logical to use the same outcome measures for all patients within one series (as in an RCT). However, since every patient is analysed separately in N-of-1 trials, the outcome measures do not have to be the same for all patients in a series. It is more important that the outcome measures are relevant for each individual patient.[5] On the other hand, in a series of N-of-1 trials, it is of interest to demonstrate variation (heterogeneity) in the results across patients. Therefore, the outcome measures for the different patients should be comparable.

Measures taken: Outcome measures were individualised.

Using the methods proposed by Guyatt, *et al* [2,6] and Beurskens, *et al* [7], all patients from series A were asked before the start of their trial to identify their most important complaints with regard to osteoarthritis. Every day each patient scored the severity of his or her own most important complaints (four to eight items) on a 7-point ordinal scale ranging between 0 = no complaints at all and 6 = unbearable complaints. Thus, each patient scored outcome measures that were relevant to him or herself. Examples of such complaints were: pain in the knee when going up and down stairs, pain in the knee when lifting shopping bags, stiffness of the knee when getting out of bed. As all items were scored on a similar scale (7-point ordinal scale), differences in the effect of the treatment on the severity of the individual complaints could be compared across patients. Outcome measures were also individualised for series B, in which the patients could select their main complaint among a set of questions regarding quantity or quality of sleep.

Results of measures:

Because we used individualised outcome measures, we were able to formulate treatment recommendations based on the results of relevant outcome measures (for the patient). In addition, the patients found it easy to answer the questions reflecting their main personal complaints, and completion rates were high. For all patients who finished their trial the completion rate for the main personal outcome measures was at least 99% in series A, and

82% to 100% in series B. Furthermore, using the same 7-point scale to score individual complaints enabled us to study heterogeneity among the patient outcomes. The results ranged from no difference to large differences in favour of non-steroidal anti-inflammatory drugs (NSAIDs) compared to paracetamol in series A, and similarly in series B from no difference to large differences in favour of temazepam compared to placebo or a lower dosage of temazepam.

Recommendations:

Identification of the main complaint in each patient, yet using the same scale to assess the outcomes for all patients in a series, enables the investigator to present each patient with individualised results and treatment recommendations and, simultaneously, to make comparisons of the results across patients.

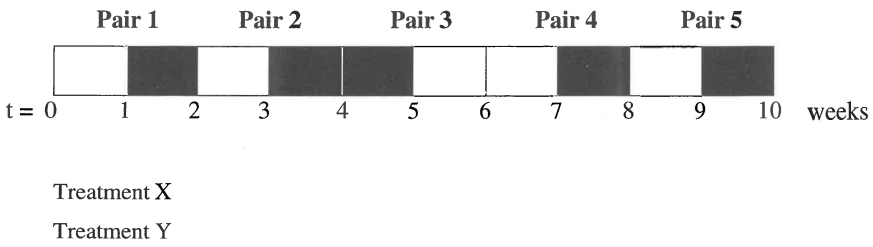

Figure 1. Example of randomisation schedule for one patient receiving 5 pairs of treatments, each consisting of one week of treatment X and one week of treatment Y

Analysis of results

Problem: How should the outcomes of the N-of-1 trials be analysed?

In both our series, the N-of-1 trials consisted of five pairs of treatment periods. Each patient received during one period of each pair the active treatment (treatment X), and

during the other period the control treatment or placebo (treatment Y). For each patient, the sequence of treatments (XY or YX) was randomised within each pair of treatment periods (Figure 1). Since N-of-1 trials are meant to evaluate the effects of treatments for each individual patient (and not to estimate the average effect in a larger population), all outcome measures should be analysed for each patient separately. The possibilities of statistical testing for significance, however, are limited, but in some studies the paired t-test has been used.[6,8]. The disadvantage of this test is that it assumes normality of the data, and independence of the data from each treatment period. The problem of autocorrelation (i.e. the data are not independent) can be solved by using the average of the data in each treatment period instead of all individual data [2], while non-parametric tests can be applied to tackle violations of the normality assumption.

Furthermore, time-series analysis may sometimes be used. In time-series analysis, data over time are compared for separate treatment periods. Time-series analysis is appropriate when there is serial dependency in the data (i.e. successively measured data are significantly correlated). One limitation is that time-series analysis requires a large amount of data from each treatment period, and various authors have stated that at least 50, but preferably 100 observations are necessary. However, relatively short treatment periods are often used in N-of-1 trials and time-series analysis is therefore not really appropriate.[9]

Also, non-parametric tests, such as the sign test or the randomisation test can be used. These tests do not require a normal distribution of the data. However, the disadvantage of the sign test is that it lacks power. At least five pairs of treatment periods are necessary before the significance level of 0.05 can be reached ($1/2^5$),[2] irrespective of the magnitude of difference between the treatments. This is the reason why we used five pairs of treatment periods in both our series.With the randomisation test, the power can be in increased by using less restricted randomisation schedules.[10,11] For example, if 5 of the 10 periods in the N-of-1 trials of our series had been randomly assigned (without restrictions) to treatment X and the other five to treatment Y, the smallest possible p-value would have been 1/252=0.004 (=1/number of randomisation possibilities, in this case 1/[10!/5!5!] = 1/252). In case of four periods of treatment X and four periods of treatment Y, this would still be 0.014 (1/[8!/4!4!]). However, restrictions are often necessary to

ensure, for example, that the active treatment periods will not all be concentrated at the end of the trial, while it could be possible that the patient is improving spontaneously.[11]

In our series, however, we did not investigate whether one treatment was better than the other, but whether one treatment was equally effective as the other. In other words, the trials we conducted were equivalence trials, rather than superiority trials, which meant that conventional significance testing could not be used for analysis.[12-14] Since 'no difference' between two different treatments cannot be proven, an equivalence range is defined before the start of the trial, i.e. a range of differences between treatments that are considered to be of no clinical importance. If the confidence interval of the difference between treatments lies entirely within this equivalence range, equivalence may be concluded with only a small probability of error.[12] To test equivalence a large number of data are required: the smaller the equivalence range, the more data that are needed.[12,15,16] However, as stated above, the number of observations are often very small in N-of-1 trials, and therefore equivalence testing is often not feasible.

Measures taken: Computation of median differences in outcomes, and definition of cut-off points.

For both series, we decided to define (a priori) cut-off points for a minimal relevant difference for each primary outcome measure. In this way we could distinguish between equivalence of effects, small, or large differences in the effectiveness of medication. Given the non-normal distribution of our data, we calculated differences in median scores between the two treatments for each outcome measure. For example, for series B, a median difference in the time to fall asleep of at least 30 minutes in favour of temazepam was considered to be a large effect, 5 to 30 minutes a small effect, and < 5 minutes no effect of temazepam. Nikles, *et al* also formulated cut-off points and defined responders, indefinite responders and non-responders according to the number of pairs in which a clear difference was seen.[17]

Results of Measures:

Although the cut-off points can be considered to be arbitrary, they enabled us to differentiate between no, small and large treatment effects, and, thus, to demonstrate variation in the results across patients.

Recommendations:

When setting up an N-of-1 trial the investigator should carefully consider the objective of the trial. If the aim is to evaluate whether another type of medication is more effective than current treatment, the trial can be designed as a superiority trial, and conventional tests of significance may be applied during the analysis. If, however, the aim is to confirm the equivalence of two (or more) treatments (which is the case in all efforts to reduce or stop medication), the N-of-1 trial is an equivalence trial, and requires a different method of analysis, with a strong emphasis on the definition of a minimal clinically relevant difference, and sufficient observations (treatment pairs) to enable equivalence testing.

Withdrawal

Problem: Withdrawal from the trials.

Six of the 13 patients (46%) in series A, and three of the 15 patients (20%) in series B, did not complete their trial period. The reasons for withdrawal (as given by the patients) are summarised in Table 2.

Table 2. Reasons for withdrawal from the trials

Reasons for withdrawal	No. of patients in series A	No. of patients in series B
Perceived lack of efficacy	4	2
Perceived side effects	1	1
Duration of trial period too long	1	-

Series A: NSAIDs or paracetamol for osteoarthritis

Series B: reduction of temazepam for sleep disturbances

Perceived lack of efficacy: Four patients in series A did not finish their trial because of perceived lack of efficacy. Three of these four patients received dosages of NSAIDs during their trial that were lower than they had taken before the trial. The other patient had taken additional paracetamol/codeine before the start of her trial, but was asked not do so during the trial. These findings suggest that the withdrawal of these patients was due to subtherapeutic dosages of medication during their trials. In series B, two patients did not complete their trial because of perceived lack of efficacy. During half of all the treatment periods in this series, the patients received either placebo or dosage of temazepam that was lower than they normally took. All patients described above were aware that they would receive lower dosages of the medication. Therefore, these patients may also have expected poor results. These expectations, in themselves, may have resulted in a perceived lack of efficacy.

Perceived side effects: In both series one patient withdrew because of perceived side effects. Before the start of the trial the patient who withdrew from series A was used to taking diclofenac tablets. To ensure blinding, however, she received capsules during the trial containing either diclofenac, and placebo, or paracetamol. In both treatment periods she reported abdominal complaints. Because only the capsules were given in both treatment periods, it is possible that the side effects were not caused by the medication, but by the gelatine capsules themselves. Similarly, the patient in series B received medication in a different dosage form (tablets instead of temazepam capsules). This patient reported nausea, and withdrew after a few days. It could also be possible that concerns regarding the effects of taking the medication in another dosage form led to reports of nausea and abdominal complaints during the trials. Finally, pure coincidence may also have resulted in reports of additional complaints during the trials.

Duration of the trial: In series A the duration of the trial was 20 weeks, and in series B it was 10 weeks, which made it easier for the patients in series B to complete their trial. In series A, one patient withdrew because the study took too long and because of coexisting complaints (back pain). None of the patients in series B withdrew because the trial took too long.

Recommendations:

A number of measures can be taken to prevent withdrawal due to subtherapeutic dosages or (perceived) side effects. Firstly, the drug can be prescribed in the same dosage and dosage form as the patient is used to. Secondly, the trial can be started with a run-in period, during which the patient receives the drug in the dosage and/or dosage form in which it will be prescribed during the trial. In this way, the patient can withdraw from the trial at an early stage, or the dosage form and/or the dosage can still be adjusted before the data-collection actually starts. Moreover, for patients who have not previously taken the medication under study, a run-in period can certainly be recommended. Finally, if blinding is difficult because of insurmountable differences in the dosage form, size, colour, smell or taste of the medication, the double-dummy design can be used. In a double-dummy design the patient receives active treatment X and placebo treatment Y during one period, and placebo treatment X and active treatment Y during the other. However, patients may withdraw because of the large number of tablets or capsules they have to take during the trial.[17]

An alternative method to prevent withdrawal due to expectations of poor results due to lower dosages is to keep the patients completely unaware of the dosages. However, this may not receive ethical approval. Furthermore, to prevent withdrawal due to the length of the trial period, this should be kept as short as possible. This may seem simple, but, in fact, this decision can be quite complex, and may affect the power of the trial (see the paragraph on analysis). Recently, in a reaction to an N-of-1 trial in a pregnant woman with nausea and morning sickness [18-20], Campbell discussed the length of the treatment periods to be completed.[21] While the total trial period should not be too long,[22] the separate treatment periods should be long enough to achieve and detect a clinically relevant treatment effect.[18]

Follow-up

Problem: reverting to the same medication as before the N-of-1 trial.

After the completion of the trials, the researcher visited the patients at their homes and discussed their personal results. Subsequently, the patients were asked about their

intentions with regard to future medication intake for osteoarthritis (series A) or sleep disturbances (series B). The general practitioners (GPs) were informed about the intentions of their patients. However, three months after the completion of series A, four of the six patients who intended to switch to paracetamol were taking NSAIDs for osteoarthritis, and one patient was taking both paracetamol and NSAIDs. In series B, two of the nine patients who intended to reduce or stop their temazepam intake had reverted to their previous dosage. The reasons for reverting to the same medication are summarised in Table 3.

Table 3. Follow-up: Reasons for reverting to the same medication as before the N-of-1 trials

Reasons for reverting in series A
• Perceived lack of efficacy of paracetamol
• Deterioration of osteoarthritis
• Misunderstanding regarding paracetamol dosage
• Preference of small tablets (diclofenac) over larger ones (paracetamol)
• Patient found it a waste to throw away the NSAIDs she still had in her possession
Reasons for reverting in series B
• Dosage of temazepam had not been adjusted after the N-of-1 trials
• Patient sees no disadvantages in taking temazepam

Series A: NSAIDs or paracetamol for osteoarthritis
Series B: reduction of temazepam for sleep disturbances

Patients were not sufficiently instructed, coached and/or monitored: One patient from series A indicated that she had reverted to taking NSAIDs because the paracetamol was not sufficiently effective. However, instead of taking 1000 mg paracetamol three times a

day, the patient had been taking 500 mg twice a day. After solving this misunderstanding, the patient agreed to start taking paracetamol again. Furthermore, another patient from series A considered it a waste to throw away the NSAIDs she still had in her possession, and wanted to finish her supply of NSAIDs before starting the paracetamol treatment. For one patient from series B the dosage of temazepam had not (yet) been reduced. Since the end of the trial, the patient had not seen the GP and neither the patient nor the GP had taken any action to adjust the treatment policy. Obviously, these three patients were not sufficiently instructed by the researcher or sufficiently coached and monitored by the GP.

Motivation to adjust treatment: In one patient from series B temazepam had little effect on the quality or quantity of sleep, but the patient saw no disadvantages in taking temazepam and decided to take it just as frequently as before the trial.

Recommendations:

To prevent patients from reverting to previous treatment, a number of measures can be taken: 1) before the start of the trial the patients should be well informed about the objectives and, especially, the consequences of the trial (i.e. continuation or adjustment of treatment policy), 2) the GP (together with the researcher) should discuss the results with the patient and adjust (if appropriate) the treatment policy immediately, 3) the patients should be well instructed and encouraged to follow the treatment recommendations, and 4) follow-up contacts by GP or researcher are recommended. Furthermore, before the start of a trial, the patient's reasons for participation should be investigated. Obviously, it is useless to start an N-of-1 trial for a patient who already knows that he/she will not change his/her treatment. Patients should participate because of their willingness to question their current treatment, and they should be willing to adjust their treatment, if necessary. Finally, the minimal relevant difference between treatments may vary per patient. Therefore, we suggest defining an individual minimal relevant difference for each patient before the start of the trial. This will establish individually relevant decision rules with clear agreements on the interpretation of results and decisions regarding the future treatment policy for each patient.

Conclusions

In conclusion, given the importance of adjusting the design of an N-of-1 trial to the specific characteristics of each individual patient, it is difficult to formulate general 'how to do it' guidelines for N-of-1 trials. When designed adequately and carried out successfully, N-of-1 trials may be of great help when deciding on drug treatment for individual patients. In our opinion, it is important to carefully consider the objective of the trial, and, with that, the randomisation schedule and the type of analysis. Furthermore, it is important to individualise the outcome measures, so that they are relevant for each patient. We recommend that the dosages and dosage forms are the same as before the trial, and that the trial starts with a run-in period. General decision rules with regard to the efficacy of the treatment may help to demonstrate variation across patients. We also recommend that individualised decision rules for the future use of medication are formulated for each patient before the start of the trial, to adjust (if appropriate) the treatment policy is adjusted immediately after the results are discussed with the patient, and finally, that the patients are provided with adequate instructions and support after this adjustment. When more experiences of N-of-1 trials have been reported, 'best practice guidelines' for N-of-1 trials can be formulated for various types of treatment problems.

Acknowledgements

This study was supported unconditionally by Leo Pharma, The Netherlands.

References

1. Altman DG. Practical statistics for medical research. London: Chapman & Hall, 1991:1-9.
2. Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D, Keller J. A clinician' s guide for conducting randomized trials in individual patients. CMAJ 1988;139:497-503.
3. Wegman ACM, Van der Windt DAWM, De Haan M, Devillé WLJM, Chin A Fo CT, De Vries ThPGM. Switching from NSAIDs to paracetamol: a series of n of 1 trials for individual patients with osteoarthritis. Ann Rheum Dis 2003;62:1156-61.
4. Wegman ACM, Van der Windt DAWM, Bongers M, Twisk JWR, Stalman WAB, De Vries ThPGM. Efficacy of temazepam in frequent users: a series of N-of-1 trials. Fam Pract. [in press]
5. Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determing optimal therapy – randomized trials in individual patients. N Engl J Med 1986;314:889-92.
6. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The n-of-1 randomized controlled trial: clinical usefulness, our three-year experience. Ann Intern Med 1990;112:293-299.
7. Beurskens AJ, De Vet HC, Koke AJ, Lindeman E, Van der Heijden GJ, Regtop W, Knipschild PG. A patient-specific approach for measuring functional status in low back pain. J Manipulative Physiol Ther 1999;22:144-8.
8. March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. N of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. BMJ 1994;309:1041-5; discussion 1045-6.
9. Kazdin AE. Statistical analyses for single-case experimental designs. In: Barlow DH, Hersen M, eds. Single case experimental designs – strategies for studying behavior change. 2nd ed. New York: Pergamon Press,1984:285-324.
10. Onghena P, Edington ES. Randomization tests for restricted alternating treatments designs. Behav Res Ther 1994;32:783-6.
11. Spiegelharter DJ. Statistical issues in studies of individual response. Scand J Gastroenterol. 1988;147:(Suppl) 40-5.
12. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996;313:36-9.
13. Senn S, Bakshi R, Ezzet N. Caution in interpretation needed. [letter] BMJ 1995;310:667.
14. Altman DG, Absence of evidence is not evidence of absence. BMJ 1995;311:485.
15. Ebutt AF, Frith L. Practical issues in equivalence trials. Statist Med 1998;1691-1701.

16. Hermens ML, Van Hout HP, Terluin B, Ader HJ, De Haan M. The importance of equivalence trials in showing the usefulness of treatments. [In Dutch: Het belang van equivalentieonderzoek om de waarde van behandelingen aan te tonen] Ned Tijdschr Geneeskd 2003;147:2162-6.
17. Nikles CJ, Gasziou PP, Del Mar CB, Duggan CM, Clavarino A, Yelland MJ. Preliminary experiences with a single-patient trials service in general practice. Med J Aust 2000;173:100-3.
18. Harker N, Montgomery A, Fahey T. Treating of nausea and vomiting during pregnancy: case presentation. BMJ 2004;328:276.
19. Harker N, Montgomery A, Fahey T. Treating nausea and vomiting during pregnancy: case progression. BMJ 2004;328:337.
20. Harker N, Montgomery A, Fahey T. Treating of nausea and vomiting during pregnancy: case outcome. BMJ 2004;328:503.
21. Campbell MJ. Commentary: statistical aspects. BMJ 2004;328:506.
22. Johannessen T, Controlled trials in single subjects: 1 Value in clinical medicine. BMJ 1991;173-4.

Chapter 6

Applicability and feasibility of N-of-1 trials in general practice: a qualitative study of GPs' views

Submitted as:
Wegman ACM, Vernooij-Dassen MJFJ, Van der Windt DAWM, De Vries ThPGM, Stalman WAB. Applicability and feasibility of N-of-1 trials in general practice: a qualitative study of GPs' views.

Abstract

Background: Although some experience has been gained with N-of-1 trials, little is known about factors that obstruct or facilitate the use of these trials in daily patient care.

Aim: To study the barriers and facilitators experienced by general practitioners (GPs) with regard to the applicability and feasibility of N-of-1 trials.

Design of study: Qualitative study based on semi-structured interviews.

Setting: General practices.

Methods: A sample of 14 GPs was purposively chosen from a group of 31 who had agreed to participate in two previous series of N-of-1 trials. The following topics regarding N-of-1 trials were investigated among GPs: 1) positive and negative experiences, 2) motivations and barriers, 3) conditions for the successful execution of N-of-1 trials, and 4) characteristics of patients who seem to be eligible for participation.

Results: N-of-1 trials were considered to be valuable and useful in general practice, because of their small scale and relevance for individual patient care. The following barriers for successfully conducting N-of-1 trials were indicated: lack of time, forgetting to carry out N-of-1 trials because they are not part of everyday patient care, and participation of patients who are not motivated. Facilitators and strategies to overcome the barriers were: willingness of the GPs to test their treatment policy, embedding the trials in the everyday routine of the practices, good research assistance, relevant trial objectives, feasible trial methods, involvement of the GPs in the development and execution of the trials, and participation of only those patients who are receptive to changes in their treatment.

Conclusions: Our findings suggest that N-of-1 trials could be integrated in daily patient care in general practice, especially if they are initiated and supported by a research group or an N-of-1 service.

Introduction

Patients should be treated according to the best evidence, which can be derived from evidence based guidelines.[1,2] However, guidelines are not strict protocols, but merely give recommendations on how to treat patients. In daily patient care, these recommendations need to be tailored to individual patients, for which N-of-1 trials can be instrumental.[3]

N-of-1 trials are carried out with one patient only. They may help a general practitioner (GP) to decide on a treatment policy when there is doubt regarding the effectiveness of medication for an individual patient. In contrast to randomised controlled trials (RCTs), N-of-1 trials do not assess the average effect for a whole population, but assess what is best for an individual patient. The patient is his or her own control, and receives during several periods of time the experimental and control treatment in random order. If possible, the patient, the physician and the researcher are all blinded for the sequence of the treatments.[3]

In addition to facilitating the tailoring of treatment to individuals, N-of-1 trials may help to increase compliance with treatment in individual patients, or may reduce the unnecessary use of certain drugs. For example, by conducting N-of-1 trials, Patel showed that treatment for 8 out of 20 patients with non-reversible chronic airflow limitation (in these cases ipratrobium bromide or theophylline) could be stopped, whereas otherwise it would have been continued.[4]

Although some experience has been gained with N-of-1 trials,[4-9] little is known about factors that obstruct or facilitate the applicability and feasibility of these trials in daily patient care. Therefore, a qualitative research method was applied to make an inventory of barriers and facilitators that general practitioners (GPs) experience with regard to the applicability and feasibility of N-of-1 trials in general practice. From the findings, we aimed to compose a checklist of conditions under which these trials may be successfully carried out. We regarded a successful trial as one where GPs did participate, and patients participated and completed their trial.

Methods

Sampling of subjects

Guided by the principles of qualitative research sampling,[10] a sample of 14 GPs was purposively chosen from a group of 31 who had agreed to participate in series of N-of-1 trials in patients with osteoarthritis (OA) of the hip or knee,[8] or sleep disturbances[9]. The sample was purposively chosen to represent a diversity of GPs with regard to the following: 1) the series for which they actually conducted N-of-1 trials (no series, one or both), 2) completion or not of the entire trial period by their patients, 3) surgery location (urban, suburban, or rural), 4) adjustment or not of treatment policy as a result of the N-of-1 trials, 5) years of work experience, 6) with or without university appointment, and 7) with or without former research experience. Two GPs were unwilling to participate, one because of the expected length of the interview, and one because of lack of interest.

Data-collection

One researcher (AW) visited the GPs and carried out semi-structured interviews. Semi-structured interviews are characterised by open-ended questions that define the area to be explored, and from which the interviewer may diverge in order to pursue an idea in more detail.[11] The GPs could discuss anything they could think of regarding the applicability and feasibility of N-of-1 trials in general practice. All interviews were carried out by the same researcher. The interviews took approximately one hour, and were audiotaped and transcribed with the aid of supplementary notes.

To identify the key barriers and facilitators with regard to the applicability and feasibility of N-of-1 trials in general practice, the following topics were discussed with the GPs during the interviews: 1) their positive and negative experiences with N-of-1 trials, 2) motivations and barriers in the execution of N-of-1 trials, 3) conditions for the successful execution of N-of-1 trials, and 4) the characteristics of patients who seem to be eligible for participation in N-of-1 trials.

Analysis

The analysis was performed by two researchers (AW, MV). In order to cluster the information obtained from the interviews, both researchers independently identified and coded each issue raised by the GPs. Subsequently, disagreements in the coding were discussed to refine the coding categories. The researchers then independently classified the identified issues into main and subcategories.[10] Again, any disagreements were discussed to refine the subcategories. Finally, all results were summarised to compose a checklist for researchers who are interested in conducting N-of-1 trials.

Results

The identified issues from the interviews were divided into three main categories: 1) positive notions or facilitators and 2) negative notions or barriers with regard to the applicability and feasibility of N-of-1 trials in general practice, and 3) possible strategies to overcome the barriers. The results of the interviews showed that each of these main categories could be subdivided into levels of 1) the GP, 2) the patient, and 3) the research.

Facilitators at GP level

At GP level, two types of facilitators were identified: 1) the relevance of the objectives of the N-of-1 trial according to the GP, and 2) the willingness of the GP to test his/her treatment policy. The last issue is illustrated by the quote below.

> Quote 1:
> *"I do a lot of things because I have been doing them for years, and they are not always evidence based, I think."*

Facilitators at patient level

Patients who are receptive to changes in their treatment were considered to be suitable candidates for N-of-1 trials. Furthermore, the following issues that motivate patients to participate were indicated: 1) the conviction that something is done about their disease or illness, and 2) their curiosity.

Facilitators at research level

N-of-1 trials were indicated as small-scale surveyable trials, which were easy to understand and conduct, and could be clearly explained to the patient. The GPs were therefore of the opinion that N-of-1 trials can be developed and carried out by the GP. Because the results of N-of-1 trials are directly applicable to their patients, the GPs considered the N-of-1 method to be useful in testing a treatment policy in individual patients. In their opinion, N-of-1 trials were a good way to convince and motivate patients to adhere more closely to their treatment if proven effective, or to discontinue treatment that demonstrated no effect.

Quote 2:

"... more support in trying to convince people."

Furthermore, the GPs indicated that a good co-operation between the GP and the pharmacist was important.

Barriers at GP level

A major barrier for the GPs to carry out N-of-1 trials is lack of time. Other barriers were lack of the necessary medical or research skills to carry out the trials, unstructured working methods, and conflict of business and patient interests.

After agreeing to carry out N-of-1 trials in patients with OA and/or patients with sleep disturbances, some GPs admitted that they forgot about the trials and therefore also forgot to select and recruit patients. The main reason for this was the fact that the trials were not part of everyday patient care, and thus not embedded in the normal routine of the general practice.

Quote 3:

"When I'm busy with a patient, I forget the trials."

Barriers at patient level

Patients who were considered by the GPs to be unsuitable for participation in an N-of-1 trial were those who are not motivated or not able to participate in the trial. The first

category includes patients who are not motivated to invest their time and energy in the N-of-1 trial and patients who are unwilling to participate in a critical appraisal of their treatment. The second category includes patients who have physical, psychological or cognitive (e.g. poor memory) impairments, as well as patients with cultural or language barriers.

Barriers at research level

It is time and energy consuming to carry out N-of-1 trials. This barrier was considered to be greater if the N-of-1 trials have to be designed by the GPs themselves. Another barrier noted by the GPs was the fact that, initially, it takes more time to conduct a trial than to treat a patient immediately. Furthermore, by definition, the results of N-of-1 trials cannot be directly translated into general decision rules regarding treatment strategies.

Strategies to overcome the barriers at GP level

As noted above, a major barrier for the GPs to carry out N-of-1 trials is lack of time. However, it was noted that delegating tasks related to N-of-1 trials to the practice assistant can save the GP time and energy.

The GPs indicated that if a research group initiates a series of N-of-1 trials, it is important that their expertise is used in the development of the trial protocols and the execution of the trials. The fact that they will not merely be data-suppliers will stimulate them to carry out N-of-1 trials. However, if their involvement is too time-consuming, this, in turn, can be a barrier.

Strategies to overcome the barriers at patient level

To increase the probability that a patient will participate and successfully complete the entire trial period, it was indicated that it is important to enrol patients who are motivated and consider the trial relevant for their personal objectives, who appreciate an objective line of reasoning, and who critically assess their treatments. The patient should also have the physical, psychological and cognitive abilities to participate.

Strategies to overcome the barriers at research level

The time and energy that GPs have to spent in N-of-1 trials can be limited by making sure that trial materials are available and ready to use, and by providing good research assistance. Furthermore, it was indicated that adequate and timely feedback from a research group about the results, including specific treatment recommendations, increases the feasibility of N-of-1 trials.

Other strategies to increase the success of N-of-1 trials

The GPs indicated that embedding N-of-1 trials in the daily routine of the practice might help to overcome several barriers simultaneously. If the trials fit in with daily practice, they will not intervene with other aspects of patient care.

> Quote 4:
> *"The general practice is busy and everything else causes turmoil and then, if they (the N-of-1 trials) are not embedded in daily practice, nothing gets done."*

Additionally, by embedding the trials in the daily routine it will take less time, and the GP may be less likely to forget about his/her activities regarding N-of-1 trials.

> Quote 5:
> *"They can easily be done during surgery hours, as far as time is concerned."*

Finally, the GPs suggested a number of general conditions that may increase the success of N-of-1 trials, including: 1) a good relationship between the GP, the patient, and the supporting researcher, 2) affiliation of the GP with the research group, 3) a trustworthy (i.e. non-profit-making) initiator of the trials, and 4) ethical and methodologically feasible trials with clear trial objectives, detailed trial manuals, and good information for both the GPs and the patients. All facilitators and barriers with regard to N-of-1 trials, and strategies to overcome the barriers are summarised in Table 1. This checklist can be used by researchers and GPs in the preparation and execution of N-of-1 trials.

Table 1. Checklist of issues that facilitate the preparation and execution of N-of-1 trials

The GP

Motivation
- The GP considers it to be important that their expertise is used in the development of the trial protocols and the execution of the trials.
- The GP is interested in the trial objectives.
- The GP considers the trial objectives relevant for patient care.
- The GP wants to test his/her treatment policy.

Time
- The GP delegates research tasks to his/her assistant.

Other
- The GP possesses all the skills needed to carry out the trials.
- The GP has structured working methods.
- The GP has no conflict of business and patient interests.
- The GP is affiliated with the research group.

The patient

Motivation
- The patient is motivated to participate in the trial.
- The patient is willing to invest time and energy in the trial.
- The patient wants to assess his/her treatment critically or needs an objective line of reasoning to be convinced of a (better) treatment policy.
- The patient considers the trial relevant for his/her personal objectives.
- The patient is receptive to changes in his/her treatment.

Abilities
- The patient has the physical, psychological and cognitive abilities to participate.
- The patient has no cultural or language barriers.

The N-of-1 trial

Ethics
- The trials are approved by an ethics review board.
- The initiator and the objectives of the trials are considered to be trustworthy (i.e. non-profit-making).
- The trials are methodologically feasible with clear trial objectives.
- Clear information is available for the GPs and the patients.

Time
- The trials can be embedded in the daily routine of the practice.
- The trials are easy to conduct.
- A research group provides good research assistance:
  * The research group provides a detailed trial manual.
  * The trial materials are available and ready to use.
  * The research group provides timely feedback about the results of the trials, including specific treatment recommendations.

Other
- The trials are used to test treatment policy in individual patients or to convince and motivate patients regarding (changes in) their treatment.
- The co-operation between GP, pharmacist, patient and researcher is satisfactory.

Discussion

The GPs described the N-of-1 trials as small-scale surveyable trials, which are easy to understand and to carry out. Because the results of N-of-1 trials are directly applicable to their patients, GPs considered the N-of-1 method to be useful for testing a treatment policy in individual patients.

A limitation of our study may be that most of the GPs had already participated in N-of-1 trials. It is therefore unknown whether other aspects of the applicability and feasibility of N-of-1 trials would have been revealed if GPs were interviewed who had not carried out N-of-1 trials.

Our study showed that, in general, the GPs considered N-of-1 trials to be feasible and applicable if they meet certain criteria at the level of the research, the GP and the patient. Using a more indirect approach, Guyatt, *et al* and Larson, *et al* also concluded that the N-of-1 method is feasible in clinical practice.[5,7] Feasibility was demonstrated by the fact that N-of-1 trials were successful in determining whether or not treatment was effective and, as a result, a treatment policy could be formulated.

We are not aware of any other studies that have systematically addressed the applicability of N-of-1 trials. However, several publications have discussed the participation of GPs and their patients in research in general. Similar to our study, these studies showed that issues such as time,[12-17] relevance of study objectives,[12-14,18,19] involvement of the GP in an early stage of the development of the study,[12,14,17] being practice-centred,[12,15,17,19,20] providing clear information,[12-15,17] and providing good research assistance[12] are important determinants of successful participation. In addition to these issues, the present study also revealed issues that are unique to N-of-1 trials. Because N-of-1 trials are used to tailor treatment to individual patients, it is very important that these patients are receptive to changes in their treatment. It would be a waste of time, effort and money to start a trial for a patient who already knows that he/she will not change his/her treatment. Therefore, patients who are willing to participate because they consider the trial relevant for their personal objectives, want to assess their treatment critically, or need an objective line of reasoning to be convinced of a better treatment policy, are especially suitable for participation in N-of-1 trials. Likewise, GPs

who are interested in testing their treatment policy in individual patients are highly suited to carry out N-of-1 trials.

Research is often considered to be extra work that is of relatively low priority, compared to patient care. N-of-1 trials, however, can be used to tailor treatment to individual patients, and can therefore be considered as part of patient care. The success of N-of-1 trials may be increased if a research group initiates the N-of-1 trials protocols while involving the GPs in the development of these protocols. In this way, they are made jointly responsible, and are thus co-owners of the trials. Co-ownership enables the GPs to determine for themselves how the N-of-1 trials can best be embedded in the everyday routine of their practice. Some GPs may find it easy to carry out N-of-1 trials during their surgery hours (see quote 5), while others may delegate some tasks related to N-of-1 trials to their practice assistant or practice nurse. For example, if a GP has doubts about treatment for a particular patient, and thinks that an N-of-1 trial could help, he or she can refer the patient to the practice assistant or nurse. Subsequently, the practice assistant/nurse explains the trial objectives to the patient and describes the course of events during the trial. Printed intake-forms are available, and can be used by the assistant/nurse to determine, in consultation with a research group and a pharmacist, the applicability and feasibility of an N-of-1 trial for this particular patient. Furthermore, the assistant/nurse can perform the trial with varying degrees of involvement of a research group and a pharmacist.

Some aspects of N-of-1 trials require a great deal of work, such as the design of questionnaires or diaries, the preparation of placebos, and the analysis of results. Therefore, support and assistance from a pharmacist and/or a research group is necessary. N-of-1 trial services, also known as single patient trial (SPT) services, have been successfully developed in Canada and the USA.[5,7] The involvement of the services provided ranges from consultants who teach clinicians how to run their own trials, to consultants who assess the value of an N-of-1 trial for a particular patient, develop the trial protocol, provide all trial materials and also perform the trial. In other words, the assistance of the N-of-1 service is tailored to the needs of the doctor.

In conclusion, our study showed that, in general, the GPs considered N-of-1 trials to be feasible and applicable if they are directly involved in the development of the trial protocols and execution of the trials, the objectives are relevant, the trial methods are

feasible, the trials are embedded in the daily routine of the practices, a research group provides good research assistance, both the GPs and the patients are motivated, and only patients who are receptive to changes in their treatment participate. Our findings suggest that N-of-1 trials could be part of daily patient care in general practice, especially if they are initiated and supported by a research group or an N-of-1 service.

Acknowledgements

We wish to thank all the general practitioners who participated in the study. This study was supported unconditionally by Leo Pharma, The Netherlands.

References

1. Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB. Guidelines. In: Evidence-based medicine: How to practice and teach EBM. 2nd ed. Edinburgh: Churchill Livingstone; 2000:169-82.
2. Rosser WW, Sharif MS. Chapter 9: Choosing guidelines with confidence. In: Evidence-based family medicine. Hamilton: BC Dekker; 1998:65-8.
3. Guyatt G, Sackett D, Adachi J et al. A clinician' s guide for conducting randomized trials in individual patients. CMAJ 1988;139:497-503.
4. Patel A, Jeaschke R, Guyatt GH, Keller JL, Newhouse MT. Clinical usefulness of n-of-1 randomized controlled trials in patients with nonreversible chronic airflow limitation. Am Rev Respir Dis 1991;144:962-4.
5. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The n-of-1 randomized controlled trial: clinical usefulness, our three-year experience. Ann Intern Med 1990;112:293-299.
6. Nikles CJ, Glasziou PP, Del Mar CB, Duggan CM, Clavarino A, Yelland MJ. Preliminary experiences with a single-patient trials service in general practice. Med J Aust 2000;173:100-3.
7. Larson EB, Ellsworth AJ, Oas J. Randomized clinical trials in single patients during a 2-year period. JAMA 1993 Dec 8;270(22):2708-12.
8. Wegman ACM, Van der Windt DAWM, De Haan M, Devillé WLJM, Chin A Fo CT, De Vries ThPGM. Switching from NSAIDs to paracetamol: a series of n of 1 trials for individual patients with osteoarthritis. Ann Rheum Dis 2003;62:1156-1161.
9. Wegman ACM, Van der Windt DAWM, Bongers M, Twisk JWR, Stalman WAB, De Vries ThPGM. Efficacy of temazepam in frequent users: a series of N-of-1 trials. Fam Pract. [in press]
10. Corbin J, Strauss A. Grounded theory research: procedures, canons, and evaluative criteria. Qualitative Sociology 1990;13:3-21.
11. Britten B. Qualitative interviews in medical research. BMJ 1995;311:251-3
12. Huibers M, van der Windt D, Boeke J. The participation of general practitioners in scientific research. [In Dutch: De deelname van huisartsen aan wetenschappelijk onderzoek.] Huisarts Wet 2002;45:454-8.
13. Kaner EFS, Haighton CA, McAvoy BR. 'So much post, so busy with practice, so no time!': a telephone survey of general practitioners' reasons for not participating in postal questionnaire surveys. Br J Gen Pract 1998;48:1067-9.

14. Ward E, King M, Lloyd M, Bower P, Friedli K. Conducting randomized trials in general practice: methodological and practical issues. Br J Gen Pract 1999;49:919-22.
15. Borgiel AEM, Dunn EV, Lamont CT, MacDonald PJ, Evensen MK, Bass MJ et al. Recruiting family physicians as participants in research. Fam Pract 1989;6:168-72.
16. Peto V, Coulter A, Bond A. Factors affecting general practitioners' recruitment of patients into a prospective study. Fam Pract 1993;10:207-11.
17. Gray RW, Woodward NJ, Carter YH. Barriers to the development of collaborative research in general practice: a qualitative study. Br J Gen Pract 2001;51:221-2.
18. Tognoni G, Alli C, Avanzini F, Bettelli G, Colombo F, Corso R et al. Randomised clinical trials in general practice: lessons from a failure. BMJ 1991;303:969-71.
19. Jonker PL, Sumajow CA. Randomised clinical trials in general practice. BMJ 1992;304:508.
20. Murphy E, Spiegal N, Kinmonth AL. 'Will you help me with my research?' Gaining access to primary care settings and subjects. Br J Gen Pract 1992;42:162-5.

Chapter 7

General Discussion

Introduction

The present thesis focuses on the evaluation of the effectiveness of drug treatment in individual patients, using N-of-1 trials in a general practice setting. Specific attention is paid to the practical and methodological difficulties encountered in N-of-1 trials, and their applicability and feasibility in general practice. In this final chapter, the main results and conclusions of this thesis are briefly summarised. Subsequently, we will discuss the implications of our results for future research, for general practice, and for patients. The chapter ends with some recommendations for future research, including suggestions for conducting N-of-1 trials.

Main findings and conclusions

Wide variation in treatment efficacy across patients

In general practice, deciding on the optimal treatment for individual patients is one of the main concerns of the general practitioner (GP). It is generally accepted that randomised controlled trials (RCTs) provide the strongest evidence for the efficacy of treatment. Our systematic review of RCTs and guidelines regarding osteoarthritis (OA) (Chapter 2) showed that the benefits of non-steroidal anti-inflammatory drugs (NSAIDs) for pain were significantly greater than those of paracetamol, but that the pooled estimate of the difference in effectiveness was small. However, NSAIDs may not be more effective than paracetamol for all patients. Preliminary subgroup analyses in some RCTs showed that NSAIDs were particularly effective for people with more severe pain. As a result, considering the fact that NSAIDs are associated with gastrointestinal complications and paracetamol is associated with very few side effects, guidelines recommend paracetamol as initial drug therapy for OA.

Our findings from the series of N-of-1 trials on the treatment of OA (Chapter 3) showed that the efficacy of NSAIDs, compared to paracetamol, varied widely across patients, ranging from no difference to large differences in favour of NSAIDs. We therefore agree with the EULAR guidelines, which explicitly emphasise the need to tailor therapy to the individual patient with OA.[1] Also, in our series of N-of-1 trials in long-term

users of the hypnotic temazepam (Chapter 4) the results varied widely across patients: from no to large effects of temazepam on the quality or quantity of sleep. Thus, in both of our series, the N-of-1 trials enabled us to determine the efficacy of drug treatment in individual patients. N-of-1 trials therefore provide information regarding decisions about drug treatment in individual patients that is complementary to that provided by RCTs (which assess mean treatment effects for a selected population).

Design of N-of-1 trials should be tailor-made

Both of our series of N-of-1 trials showed that N-of-1 trials may facilitate difficult decisions concerning drug treatment in individual patients. However, conducting N-of-1 trials is not as simple as it may seem. Indeed, before, during, and after data-collection, we encountered a number of difficulties, especially with regard to outcome assessment, analysis of the results, the withdrawal of patients, and follow-up. In Chapter 5 we discussed our solutions to these difficulties. In our opinion, it is important to individualise outcome measures, and to carefully consider the objective, the type of randomisation and the analysis. Furthermore, we recommend that the patient is given the same dosages and dosage forms as before the trial, and that the trial is started with a run-in period. We also recommend that both general and individualised decision rules are formulated with regard to the efficacy of the treatment, that the treatment policies are adjusted immediately after the trial, and that the patient is provided with adequate instructions and support if the treatment is adjusted. In conclusion, N-of-1 trials may be carried out successfully, provided that the design is tailored to the specific characteristics of each individual patient and the underlying medical problem (e.g. individualised outcome measures, trial medication).

GP, patient and research characteristics determine the applicability and feasibility of each N-of-1 trial

As stated above, N-of-1 trials are considered to be a feasible way in which to tailor treatment to individual patients from a research point of view. However, to be really useful for the GP, they should be applicable and feasible in general practice. Our qualitative study (Chapter 6) showed that GPs considered N-of-1 trials to be valuable and useful for general practice, because of their small scale and relevance for individual

patient care. Since the results of an N-of-1 trial are directly applicable to a patient, they considered the N-of-1 trials to be useful for testing treatment policy in individual patients. However, we found that the applicability and feasibility of each N-of-1 trial depend on several criteria at research, GP and patient level. The GPs indicated that lack of time, doubts about the relevance of trial objectives, functioning as data-suppliers only, and the participation of patients who are not motivated can be barriers to the success of conducting N-of-1 trials. Facilitators and strategies to overcome these barriers were also indicated at patient, GP, and research level. In general, the GPs considered that N-of-1 trials are feasible and applicable if only those patients participate who are receptive to changes in their treatment, the objectives are relevant, the trial methods are feasible, the GPs are willing to test their treatment policy, the trials are embedded in the daily routine of the practice, and a research group provides good research assistance. These findings suggest that N-of-1 trials could be integrated in daily patient care in general practice, especially if they are supported by a research group.

Implications for future research, general practice, and patients

Implications for future research

Pooling the results of N-of-1 trials

It could be debated whether or not to pool the results of N-of-1 trials, i.e. to quantitatively summarise the results the regarding efficacy of treatments. In our opinion, the results of N-of-1 trials should not be statistically pooled. Our series of N-of-1 trials have shown that there is a wide variation in drug efficacy (from no to large treatment effects) across patients. Similar to systematic reviews, it is pointless to pool heterogeneous results in order to calculate the 'mean' efficacy of a drug. However, for some types of drugs and medical problems the efficacy of drug treatment may show a negligible variation between patients, and pooling may seem sensible. In such cases, however, N-of-1 trials are not necessary, unless patients need to be convinced of the existence (or absence) of treatment effects, before they are willing to start (or discontinue) treatment with the specific drug.

The results of a series of N-of-1 trials may generate hypotheses regarding greater or smaller treatment effects for subgroups of patients. However, in such cases, an RCT is a more efficient and a more correct method with which to test these hypotheses, rather than performing subgroup analyses of the pooled results of N-of-1 trials. In RCTs, there is a more strict selection of patients than in N-of-1 trials. In addition, the outcome measures in N-of-1 trials can be individualised, whereas they are exactly the same for each patient in an RCT. Therefore, the findings with regard to the efficacy of treatment in different subgroups may be less biased in RCTs than in N-of-1 trials.

Treatment questions that may be tackled by N-of-1 trials

In our series of N-of-1 trials we investigated whether one treatment was equally effective as the other (OA series), or whether there was any treatment effect at all (sleep disturbances series). Other treatment questions that may be tackled by N-of-1 trials are summarised in Box 1. These treatment questions were derived from our semi-structured interviews with GPs regarding the applicability and feasibility of N-of-1 trials in general practice, and have not been presented elsewhere in this thesis.

Box 1. Treatment questions that may be tackled by N-of-1 trials

- Is the drug (still) effective?
- Is the drug more effective than another?
- Is the drug sooner effective than another?
- Is a safer drug equally effective?
- Is a cheaper drug equally effective?
- Is one drug accompanied with less (serious) side effects than another?
- Are perceived side effects really caused by the drug?
- Which dosage (of a certain drug) is best?
- Which dosage form (of a certain drug) is best?
- What time is best to take the drug?

N-of-1 trials versus usual care (i.e. trial and error)

Normally, when a GP has any doubts concerning the applicability of treatment recommendations (derived from RCTs) to a specific patient, the trial and error method is applied. This means that a particular drug will be prescribed, and will subsequently be continued if considered effective, or changed if considered not beneficial. An N-of-1 trial, unlike the trial and error method, is an evidence based method to determine which treatment is best for an individual. Therefore, it is reasonable to argue that with the help of N-of-1 trials GPs are able to give patients care that is better than the usual care (i.e. better than with the trial and error method).

However, it is impossible and undesirable to tackle each treatment problem with an N-of-1 trial. First, there should be considerable doubts about the treatment policy; secondly, the disease, the disorder, or the complaint has to be chronic or recurrent, or drugs will be prescribed for a long period of time or for frequently repeated periods of time; and thirdly, the treatment effects of the interventions should have a rapid onset and stop acting soon after discontinuation.[2]

To determine whether N-of-1 trials can indeed, lead to care that is better than the usual care, RCTs may be carried out for various medical problems with different kinds of characteristics. The objective of such an RCT would be to study whether N-of-1 trials lead to better (more effective or safer) treatment policies than those achieved by the trial and error method. Mahon and colleagues have performed a randomised study, comparing N-of-1 trials to standard practice with regard to theophylline for patients with irreversible chronic airflow limitation (CAL).[3] In this study, guiding theophyline therapy by N-of-1 trials did not improve the quality of life or exercise capacity, or reduce the overall theophylline usage during a period of 1 year, compared to standard practice. However, methodological problems with the N-of-1 trials may have negatively influenced the results of the N-of-1 trials. Nearly 50% (16/34) of the N-of-1 trials were based on only two or less pairs of active and placebo treatment periods. Furthermore, before randomisation, patients who were sure that theophylline was helpful following open therapy were excluded, whereas, in our opinion, conducting N-of-1 trials in those patients may be particularly interesting. Therefore, this approach has limited the generalisability of the results.[3]

In our opinion, N-of-1 trials may be most advantageous, compared to the trial and error method, if the treatment effects are likely to be influenced by the beliefs of the patient or the GP, thus, if (1) a patient needs to be convinced of the absence of treatment effects (i.e. placebo-effect) in order to discontinue treatment, or (2) a patient (or GP) needs to be convinced of the existence of a treatment effect in order to start, continue or adjust the treatment policy. Since the minimal relevant difference between treatments may be different for each individual patient, we emphasise that an individual minimal relevant difference should be defined for each patient before the start of the N-of-1 trial. In this way, individually relevant decision rules can be formulated with clear agreements about the interpretation of the results and decisions regarding the future treatment policy, and N-of-1 trials can be most effective.

N-of-1 trials may not have many advantages over the trial and error method if a patient has recently been diagnosed, has not yet been prescribed any treatment, and neither the patient nor the GP are prejudiced with regard to the treatment possibilities. However, if in such cases the trial and error method is used, it is still questionable whether the patient will be prescribed the 'best' treatment. For example, the patient may be quite satisfied with the first treatment that is offered and, as a result, this treatment is continued. As a consequence, the alternative treatment, although possibly better for the patient, will not be offered at all. In addition, fluctuations in the natural course of the disease may bias the perceived treatment effects, which may result in the prescription of a non-optimal type of treatment.

It could be argued that conducting N-of-1 trials is more expensive than providing usual care. However, usual care may be accompanied by unnecessary costs, such as unnecessary drug use, or the prescription of expensive drugs. In addition, the drugs or dosage forms that are prescribed may not be the most favourable for the patient, or non-optimal dosage schedules may be adhered to. This may result in lack of efficacy, side effects, and thus a higher utilisation of health care. Therefore, it would be useful to combine RCTs comparing N-of-1 trials to usual care (as described above) with a cost-effectiveness analysis. Pope, *et al* undertook such a study, and investigated whether conducting N-of-1 trials with diclofenac/misoprostol (Arthrotec) was superior and cost-effective, compared to standard treatment for OA.[4] Patients who were not certain that 'their' NSAID was helpful, or patients who were unexposed to NSAIDs, were

randomised either to participate in an N-of-1 trial or to receive conventional treatment. The patients in the conventional treatment group were asked to stop taking their NSAID, and to wait and see what happened. These patients received standard care, which could include, if necessary, high doses of paracetamol, switching strategies to NSAIDs, or other treatment such as intraarticular steroids, physiotherapy and occupational therapy. In the N-of-1 trial group, diclofenac/misoprostol (Arthrotec) was compared with placebo. After 3 months, both groups received standard care. Scores on HAQ pain and disability,[5] general health (SF-36),[6] and patient global assessments (100 mm visual analogue scale, VAS) improved more in the N-of-1 trial group, but were not statistically significant, possibly due to the small number of patients. At 6 months, 81% of the participants in the N-of-1 trial group and 79% of the patients receiving conventional treatment were taking NSAIDs. The costs of N-of-1 trials were higher than conventional treatment, due to the costs of formulating a placebo, the time spent by the nurse and the physician, and the time spent by the patient completing efficacy questionnaires or other outcome measures.

It would have been interesting if, in the N-of-1 trial group, NSAIDs had been compared to paracetamol instead of placebo, because many patients with OA need some kind of pain medication. This could have resulted in even better health outcomes, and could also have reduced the costs for participants in the N-of-1 trial group. Nevertheless, the additional research costs of N-of-1 trials may often outweigh the costs of conventional treatment. Consequently, the health benefits (effects, side effects) of N-of-1 trials need to be sufficiently important enough to result in cost-effectiveness ratios that are acceptable or in favour N-of-1 trials. It is not unlikely that such beneficial results of N-of-1 trials will need a long-term assessment of effects, side effects, and costs, especially since N-of-1 trials are applicable for patients with chronic conditions.

Implications for general practice

Embedding N-of-1 trials in general practice

Our results showed that N-of-1 trials may be useful in general practice, especially if they are supported by a research group and embedded in the daily routine of the general practice. To facilitate the implementation of N-of-1 trials in daily patient care, it should

be made clear to GPs that an N-of-1 trial should not be considered as a *trial*, but as an *additional test* that can be requested just like any other test, such as a laboratory test. If a GP requests an N-of-1 trial, it may be carried out by a research group, and the research group will send the results (including treatment recommendations) to the GP. The GPs who participated in our applicability study indicated that they wanted to be involved in the design and execution of an N-of-1 trial to a greater or lesser extent. Therefore, research groups (or N-of-1 services) are probably most valuable if their support is tailored to the needs of the GP, ranging from consultants who teach GPs how to run their own trials, to consultants who assess the value of an N-of-1 trial for a particular patient, develop the trial protocol, provide all trial materials (trial medication, questionnaires) and perform the trial. N-of-1 trial services (N-of-1 services), also known as single patient trial (SPT) services, have already been developed in Canada and the USA.[7,8] In Australia an N-of-1 service has also been developed, and approximately 175 GPs have indicated that they are interested in utilising this service. After media publicity, the N-of-1 service received over 100 phone calls from potential trial participants, demonstrating considerable consumer demand.[9]

Education in N-of-1 trials

A prerequisite for the implementation of N-of-1 trials in general practice is that GPs are familiar with these trials. This can be achieved by educating medical students, GP-trainees, and GPs about N-of-1 trials, preferably by teaching them how to (partially) design and conduct an N-of-1 trial. GPs can be educated during the course of their continuing medical education. Moreover, it would be worthwhile to train the practice nurse or assistant to carry out specific parts of the N-of-1 trials. Special attention should be paid to the fact that both GPs and medical students learn that even in uncertain situations they can prescribe drugs in an evidence based way. An additional advantage of N-of-1 trials is that this is a small-scale method and therefore suitable for medical students to become acquainted with research, from designing the trial to carrying out the analysis, in a relatively short period of time. A nice example of N-of-1 trials carried out by a medical student was descriped by Ham and Assendelft.[10]

Costs

The preparation and execution of N-of-1 trials involves costs. Because these costs may vary across different kinds of medical problems (trial medication, duration of the trial), the number of patients, the number of other people that are involved (GP, practice nurse, pharmacist, researchers, staff from an N-of-1 service) and their salary, and the country in which the N-of-1 trials take place, we are not able to present an accurate estimate of N-of-1 trial costs. However, to give some indication of the costs involved, Table 1 shows what should be taken into account when estimating the costs of an N-of-1 trial. It should be emphasised that preparing and conducting N-of-1 trials for a number of patients with the same type of medication and outcome measures (although individualised), may reduce the expenses to a great extent.

Furthermore, to be able to implement N-of-1 trials successfully in general practice, it should be clear who is going to bear these costs. As stated above, an N-of-1 trial may be considered as an additional test that can be requested just like any other test, such as a laboratory test. As such, N-of-1 trials could be considered to be part of patient care, and it could be argued that health insurance companies and/or the government should provide the finances if cost-effectiveness has been established.

Implications for patients

Opinion of patients with regard to N-of-1 trials

We have discussed the methodology, the feasibility and the applicability of N-of-1 trials, and their implications for future research, general practice and education. However, one important aspect has yet not addressed: the opinion of patients with regard to N-of-1 trials.

If patients participate in N-of-1 trials, it is investigated which treatment is best for each of them individually. In most instances, the patients would probably appreciate this. After a series of N-of-1 trials carried out by Nikles, *et al*, the patients gave very positive feedback with regard to the N-of-1 trials.[11] However, participating in N-of-1 trials may present some difficulties for the patients. Patients have to spend time and effort on the N-of-1 trials. It is not known whether, or to what extent, patients are prepared to do that.

Table 1. What to take into account when estimating the costs of a single N-of-1 trial

Personnel	
Designing forms for patient information and exploratory interview	Researcher or GP
Exploratory interview (information and consent)	Researcher, GP, or Practice Assistant or Nurse
Designing diaries	Researcher or GP
Developing randomisation schedule and producing trial medication	Pharmacist
Appointments with the patient during the N-of-1 trial: 1) to collect completed diaries and left-over trial medication 2) to hand out diaries and trial medication for the following study period, 3) to collect, if necessary, information on additional outcome measures	Researcher, GP, or Practice Assistant or Nurse
Travelling expenses, if appropriate	Patient, Researcher, GP, and/or Practice Assistant or Nurse
Data entry	Researcher, or Practice Assistant or Nurse
Data analysis and formulation of recommendations	Researcher or GP
Final visit to discuss results and recommendations with the patient	Researcher, GP, or Practice Assistant or Nurse
Costs of materials	
Printing of 1) the explanatory interview form, 2) written information for the patient, 3) diaries, forms for additional outcome measures.	
Production and storage of trial medication	
Medication boxes	
Materials for function tests, laboratory tests, and the like, if necessary	

Furthermore, the patients may not be willing to accept the consequences of the results of the N-of-1 trials (i.e. to start, (dis)continue, or adjust the treatment). Our series in long-term users of temazepam showed that the reason or motivation for the patient to participate may be very important. Patients who participated in the trial because they wanted to see if they could reduce their temazepam intake were more successful in reducing their temazepam use than patients who participated for the benefit of science. Therefore, before the start of an N-of-1 trial, both patient and GP should agree on individualised decision rules with regard to future treatment, based on the results of the N-of-1 trial. In future research, it would be interesting to interview patients, like we interviewed GPs, and to discuss their opinions concerning the benefits, motivations, and barriers with regard to participating in N-of-1 trials.

N-of-1 trials may lead to shared decision making and patient empowerment

In addition to helping to tailor treatment to individual patients, N-of-1 trials may lead to shared decision making, and thus to patient empowerment, provided that the patients are actively involved in defining individualised decision rules with regard to the drug treatment. Shared decision making is characterised as follows: 1) at least two participants, the physician and the patient, are involved, 2) both the physician and patient take steps to participate in the decision making process, 3) sharing information is a prerequisite for shared decision making, and 4) a treatment decision (which may even be to do nothing) is made, and both parties agree with the decision.[12]

Shared decision making is in accordance with the sub-category of the Dutch Civil Code regarding the provisions applicable to the contract for medical treatment, the Contracts (Medical Treatment) Act (in Dutch: Wet op de Geneeskundige Behandelingsovereenkomst [WGBO]), as decreed in 1995. According to this Act, the physician has to inform the patient about tests, treatment options, and possible side effects. Subsequently, the physician and the patient decide together what to do.[13] In a study of patients' expectations with regard to care during surgery visits, it was found that it was important for patients to discuss their own ideas about how to manage their condition.[14] Their awareness of different treatment options and their (dis)advantages,[15] and, probably, their participation in the formulation of decision rules with regard to drug

treatment, can be beneficial for compliance with treatment and may lead therefore to better health care.

Elwyn has studied the concept of shared decision making, and proposed a model of clinical decision making processes (Figure 1).[16] The model illustrates that shared decision making can be seen as one (the best) of the following possibilities in decision making processes regarding treatment for patients: (1) the GP makes use of relevant information and makes the treatment decision (paternalism), (2) the patient makes use of relevant information (provided by the GP or elsewhere) and makes the treatment decision (informed choice), (3) the patient makes no use of relevant information that the GP can provide and dominates the decision making process (consumerism), (4) the patient is well-informed (maybe even better informed than the GP), but the GP does not take the view of the patient into account (professional dominance), and (5) both GP and patient are well-informed and decide together (who decides) what to do (shared decision making).

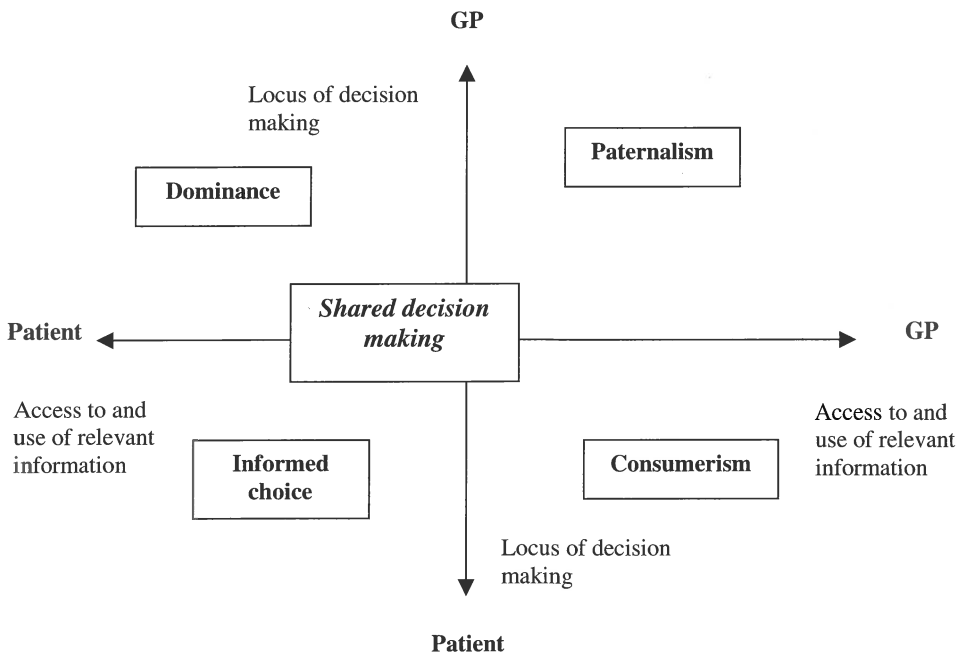

Figure 1. A model of decision making in consultations[16]

Recently, a large national survey among 195 GPs in The Netherlands has shown that, since the introduction of the Contracts (Medical Treatment) Act, shared decision making is becoming more common in general practice.[13,17] N-of-1 trials may facilitate shared decision making, especially in situations in which either the GP or the patient has to be convinced of the need to change or to adjust the treatment.

Recommendations for future research

Throughout this thesis we have given different kinds of recommendations with regard to the preparation and execution of N-of-1 trials. For easy reference, these recommendations, and also the general conditions for conducting N-of-1 trials proposed by Guyatt, *et al*, are summarised in Table 2.[2]

Furthermore, like all research, this thesis also leads to questions that need to be answered. First, it is important to identify the medical problems for which N-of-1 trials, indeed lead to better (more effective or safer) treatment policies than would be achieved by usual care would. Therefore, RCTs comparing N-of-1 trials to usual care should be carried out for various medical problems with different kinds of characteristics. In addition, it would be useful to include a cost-effectiveness analysis in such RCTs.

Secondly, it would be worthwhile to study in more detail how N-of-1 trials could be implemented most successfully in general practice. In this context, it is important to assess what kind of support could best be provided by N-of-1 services, and therefore, which professionals (researcher, GP, pharmacist, others) should be represented in an N-of-1 service. It would also be valuable to interview patients and to study their opinion concerning the benefits, motivations, and barriers with regard to participating in N-of-1 trials. Finally, it is useful to investigate whether N-of-1 trials could be feasible and applicable in secondary or tertiary care.

Table 2. Factors that facilitate the preparation and execution of N-of-1 trials in general practice

General conditions to conduct N-of-1 trials [2]
There is considerable doubt regarding treatment policy
The disease, disorder, or complaint is chronic or recurrent, or drugs will be prescribed for a long period of time or for frequently repeated periods of time
The effects of the interventions have a rapid onset and stop acting soon after discontinuation
An optimal duration of treatment periods is feasible
Clinically relevant targets are measured
Sensible criteria for stopping the trial are established
Strategies for interpreting the data are appropriate
The patient understands the trial and is enthusiastic about participating
A pharmacist is involved
The N-of-1 trial meets medical ethics criteria
Treatment questions that may be tackled by N-of-1 trials
Is the drug (still) effective?
Is the drug more effective than another?
Is the drug sooner effective than another?
Is a safer drug equally effective?
Is a cheaper drug equally effective?
Is one drug accompanied with less (serious) side effects than another?
Are perceived side effects really caused by the drug?
Which dosage (of a certain drug) is best?
Which dosage form (of a certain drug) is best?
What time is best to take the drug?

Table 2. Continued

Methodological recommendations
The objective of the N-of-1 trial is carefully considered: equivalence or superiority of treatment
The trial starts with a run-in period
If possible, dosages and dosage forms that are the same as before the trial, are given to the patient
The type of randomisation and, with that, the analysis of results are carefully considered
Before the start of the trial, general cut-off points with regard to the efficacy of the treatment are formulated to help demonstrating variation across patients
The outcome measures are individualised
Before the start of the trial, individualised decision rules with regard to the future use of medication are formulated
Practical recommendations
The GP is familiar with N-of-1 trials (objective, relevance)
The GP considers N-of-1 trials relevant for patient care and is willing to test his/her treatment policy
Patients who participate are receptive to changes in their treatment policy
The patient has the physical, psychological, and cognitive abilities to participate
There are no cultural or language barriers for the patient to participate
The N-of-1 trials are embedded in the daily routine of the practice
The practice nurse is trained to carry out specific parts of the N-of-1 trials tailored to the practice needs
The research group (N-of-1 service) provides support tailored to the practice needs
The research group (N-of-1 service) provides timely feedback about the results of the trials, including specific treatment recommendations
If appropriate, the GP adjusts the treatment policy immediately after discussing the results with the patient
Following treatment adjustments, adequate instructions and support are provided by the GP and/or research group (N-of-1 service)
It is clear who is going to bear the costs

References

1. Pendleton A, Arden N, Dougados M, Doherty M, Bannwarth B, Bijlsma JW et al. EULAR recommendations for the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis 2000;59:936-44.
2. Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D, Keller J. A clinician' s guide for conducting randomized trials in individual patients. CMAJ 1988;139:497-503.
3. Mahon Jl, Laupacis A, Hodder RV, McKim DA, Paterson NA, Wood TE, Donner A. Theophylline for irreversible chronic airflow limitation: a randomized study comparing n of 1 trials to standard practice. Chest 1999;115:38-48.
4. Pope JE, Prashker M, Anderson J. The efficacy and cost effectiveness of N of 1 studies with diclofenac compared to standard treatment with nonsteroidal antiinflammatory drugs in osteoarthritis. J Rheumatol 2004;31:140-9.
5. Fries JF, Spitz P, Kraines G et al. Measurements of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
6. Jenkinson C, Coulter A, Wright L. Short Form 36 (SF-36) health survey questionnaire: Normative data for adults of working age. BMJ 1993;306:1437-40.
7. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The n-of-1 randomized controlled trial: clinical usefulness, our three-year experience. Ann Intern Med 1990;112:293-299.
8. Larson EB, Ellsworth AJ, Oas J. Randomized clinical trials in single patients during a 2-year period. JAMA 1993 Dec 8;270:2708-12.
9. Nikles CJ, Glasziou PP, Del Mar CB, Duggan CM, Mitchell G. N-of-1 trials - practical tools for medication management. Aust Fam Phys 2000;29:1108-12 .
10. Ham MF, Assendelft WJJ. Feasibility of N-of-1 trials. [In Dutch: De haalbaarheid van N=1-onderzoek.] Huisarts Wet 2001;44:204-6.
11. Nikles CJ, Glasziou PP, Del Mar CB, Duggan CM, Clavarino A, Yelland MJ. Preliminary experiences with a single-patient trials service in general practice. Med J Aust 2000;173:100-3.
12. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (Or it takes at least two to tango). Socal Science and Medicine 1997;44:681-92.
13. Van den Brink-Muinen A, Van Dulmen AM, De Haes JC, Visser A, Schellevis FG, Bensing JM. The involvement of patients in the process of decision making. [In Dutch: De betrokkenheid van patiënten bij het besluitvormingsproces.] In: Van den Brink-Muinen A,

Van Dulmen AM, Schellevis FG, Bensing JM (reds). Tweede nationale studie naar ziekten en verrichtingen in de huisarstpraktijk. Oog voor communicatie: huisarts-patiënt communicatie in Nederland. Utrecht: NIVEL, 2004:109-20.

14. Kravitz RL, Cope DW, Bhrany V, Leake B. Internal medicine patients' expectations for care during office visits. J Gen Intern Med 1994;9:75-81.
15. Van Dulmen AM, Bensing JM. The effort of context in healthcare. A programming study. Utrecht: NIVEL, 2001.
16. Elwyn G. Discussion: conclusions and recommendations. In: Elwyn G. Shared decision making – patient involvement in clinical practice. Wageningen: Ponsen & Looijen B.V.,2001:165-88 [Thesis].
17. Schellevis FG, Westert GP, De Bakker DH, Groenewegen PP. Other methodological aspects. [In Dutch: Overige methodologische aspecten.] In: Van den Brink-Muinen A, Van Dulmen AM, Schellevis FG, Bensing JM (reds). Tweede nationale studie naar ziekten en verrichtingen in de huisarstpraktijk. Utrecht: NIVEL, 2004:123-37.

Summary

In general practice, deciding on the optimal treatment for individual patients is one of the main concerns of the general practitioner (GP). It is generally accepted that randomised controlled trials (RCTs) provide the strongest evidence for the efficacy of treatment. However, there may be reasonable doubt about whether the results of RCTs can be generalised to individual patients in general practice.

If the GP or the patient doubts the effectiveness and/or safety of a certain type of drug treatment, and prefers an objective evaluation, the N-of-1 randomised controlled trial (N-of-1 trial) may, in certain situations be a valuable addition to RCTs and a useful tool in general practice. The N-of-1 trial is a multiple crossover trial in one patient, that consists of a series of pairs of treatment periods. The patient is his or her own control, and receives during one period of each pair the active treatment, and during the other period the control intervention or placebo. The sequence of treatments is randomised within each pair of treatment periods. The patient, the physician, and the researcher are all blinded for the sequence of treatments.

The first aim of this thesis was to study whether the results of N-of-1 trials can be complementary to evidence derived from RCTs for decisions regarding drug treatment for individual patients. We therefore performed a systematic review of RCTs and guidelines concerning drug treatment for osteoarthritis (OA) of the hip or knee ***(Chapter 2)***, a series of N-of-1 trials in patients with OA of the hip or knee ***(Chapter 3)***, and a series of N-of-1 trials in patients with sleep disturbances ***(Chapter 4)***. Both series of N-of-1 trials were performed in general practice.

In our systematic review ***(Chapter 2)*** the available evidence on the efficacy of non-steroidal anti-inflammatory drugs (NSAIDs), compared to paracetamol, was summarised, and the quality and content of clinical guidelines for drug treatment of OA was compared. Published reports of RCTs and clinical guidelines were identified by a systematic search of bibliographic databases and relevant websites. In addition, the references of all retrieved articles, including systematic reviews, were screened for potentially relevant publications. The search strategy resulted in the identification of 5 RCTs and 9 guidelines. Statistical pooling of the data from 3 RCTs with adequate methodological quality and sufficient data presentation resulted in a pooled standardised

mean difference for general pain or pain at rest of 0.33 (95% CI 0.15 to 0.51), indicating a small effect in favour of NSAIDs. Pooled estimates for overall change, pain on motion and functional disability were smaller, and showed borderline significance in favour of NSAIDs. Considering the fact that NSAIDs are only slightly more effective than paracetamol and are associated with gastrointestinal complications, while paracetamol is associated with very few side effects, guidelines recommend paracetamol as initial drug therapy for OA.

In addition to the systematic review, in which average treatment effects were studied, we performed two series of N-of-1 trials to evaluate treatment effects for individual patients. The first series is described in ***Chapter 3***. In this series it was investigated, for individual patients who had been taking NSAIDs regularly, whether paracetamol is as effective as NSAIDs in the treatment of pain and disability related to OA of the hip or knee. Outcomes were assessed by daily diaries and during home visits made by the investigator every two weeks, at the end of each treatment period. Primary outcome measures were *(a)* severity of the individual main complaints and *(b)* intensity of the pain. Secondary outcome measures were *(a)* satisfaction with medication and *(b)* function test (walking 2.5 metres to and fro, sitting down and standing up five times, or going up and down stairs [time in seconds]). Every two weeks, the patient performed the same function test. Furthermore, every two weeks, information was collected on the occurrence of side effects and co-interventions (including concomitant drugs and other conservative methods of treatments). Our findings showed that the efficacy of NSAIDs, compared to paracetamol, varied widely across patients, ranging from no difference to large differences in favour of NSAIDs. N-of-1 trials can therefore be used to investigate which treatment is best for a specific person, thus avoiding unnecessary prolonged treatment with NSAIDs. However, our study also showed that there may be practical reasons why patients decide whether or not to switch from NSAIDs to paracetamol.

The second series of N-of-1 trials is described in ***Chapter 4***. There is considerable doubt about whether benzodiazepines are still effective for sleep disturbances when taken for a prolonged period of time. In addition, the use of benzodiazepines is associated with adverse cognitive effects and an increased risk of motor vehicle accidents, falls, and fractures. Therefore, in practice guidelines for the management of insomnia it is recommended that the long-term use of benzodiazepines should be prevented or stopped.

However, attempts made by GPs and their patients to stop long-term use of benzodiazepines are often unsuccessful. Probably the most important obstacle for success is the patient's awareness of the abstinence from his or her daily dose of benzodiazepines. N-of-1 trials may provide more insight into the question whether the use of hypnotics can be successfully reduced in individual patients without the loss of quality and quantity of sleep. In our series of N-of-1 trials we investigated for individual long-term users of temazepam whether placebo was as effective as temazepam, or whether 10 mg was as effective as 20 mg of temazepam, and whether these results influenced their future temazepam use. The main outcome measures were the individual main complaint and the time to fall asleep. The results varied widely across patients: from no to large effects of temazepam on quality or quantity of sleep. Our results showed that N-of-1 trials seem to be valuable for patients who are motivated to stop or reduce their temazepam use. They clearly demonstrate the (in)efficacy of temazepam, and may give patients additional confidence to discontinue taking hypnotics regularly. The value of N-of-1 trials for patients who are less motivated is unclear, because the magnitude of the treatment effect did not seem to influence their future use of hypnotics.

Thus, in both of our series, the N-of-1 trials enabled us to determine the efficacy of drug treatment in individual patients. N-of-1 trials therefore provide information concerning decisions with regard to drug treatment in individual patients that is complementary to the results of RCTs (which demonstrate mean treatment effects for a selected population).

The second aim of this thesis was to study the practical and methodological difficulties encountered in N-of-1 trials. Before, during, and after data-collection, difficulties regarding outcome assessment, analysis of the results, the withdrawal of patients, and follow-up had to be dealt with. In ***Chapter 5***, we discussed our solutions to these difficulties. In our opinion, it is important to individualise the outcome measures, and to carefully consider the objective, the type of randomisation and the analysis. Furthermore, we recommend that the patient is given the same dosages and dosage forms as before the trial, and that the trial is started with a run-in period. We also recommend that both general and individualised decision rules are formulated with regard to the efficacy of the treatment, that the treatment policies are adjusted immediately after the trial, and that the

patient is provided with adequate instructions and support if the treatment is adjusted. In conclusion, N-of-1 trials may be carried out successfully, provided that the design is tailored to the specific characteristics of each individual patient and the underlying medical problem (e.g. individualised outcome measures, trial medication).

The third aim of this thesis was to study the applicability and feasibility of N-of-1 trials in the treatment of individual patients in general practice. We performed a qualitative study to make an inventory of barriers and facilitators with regard to the applicability and feasibility of N-of-1 trials, as experienced by GPs ***(Chapter 6)***. A sample of 14 GPs was purposively chosen from a group of 31 who had agreed to participate in one or both series of N-of-1 trials described in Chapters 3 and 4. During semi-structured interviews, the following topics of N-of-1 trials were investigated among the GPs: 1) their positive and negative experiences with N-of-1 trials, 2) motivations and barriers in the execution of N-of-1 trials, 3) conditions for the successful execution of N-of-1 trials, and 4) the characteristics of patients who seem to be eligible for participation in N-of-1 trials. The GPs in our study considered N-of-1 trials to be valuable and useful for general practice, because of their small scale and relevance for individual patient care. Since the results of an N-of-1 trial are directly applicable to the patient, they considered the N-of-1 trials to be useful for testing treatment policy in individual patients. However, we found that the applicability and feasibility of each N-of-1 trial depend on several criteria at research, GP and patient level. The GPs indicated that lack of time, forgetting to carry out N-of-1 trials because they are not part of everyday patient care, and the participation of patients who are not motivated, can be barriers to the success of N-of-1 trials. Facilitators and strategies to overcome these barriers were also indicated. In general, the GPs considered N-of-1 trials to be feasible and applicable if the GP is willing to test the treatment policy, the trials are embedded in the daily routine of the practice, a research group provides good research assistance, the objectives are relevant, the trial methods are feasible, the GP is directly involved in the development of the trial protocols and the execution of the trials, and only patients who are receptive to changes in their treatment participate. These findings suggest that N-of-1 trials could be integrated in daily patient care in general practice, especially if they are initiated and supported by a research group or N-of-1 service.

Finally, in ***Chapter 7***, the main results and conclusions of this thesis are briefly summarised. Subsequently, attention is paid to the implications of the results of this thesis for future research. Since our series of N-of-1 trials have shown that there is a wide variation in drug efficacy (from no to large treatment effects) across patients, it was argued not to statistically pool the results of N-of-1 trials. Then, treatment questions that could be tackled by N-of-1 trials were indicated. Furthermore, it was recommended that RCTs should be carried out to compare N-of-1 trials with usual care for various medical problems with different kinds of characteristics. The objectives of such RCTs would be to study whether N-of-1 trials lead to better (more effective or safer) treatment policies and whether N-of-1 trials are cost-effective, compared to usual care.

Next, attention is paid to the implications of the results for general practice. The implementation of N-of-1 trials in general practice may be facilitated if (1) GPs consider N-of-1 trials as an additional test that can be requested just like any other test, such as a laboratory test, (2) research groups (or N-of-1 services) provide support that is tailored to the needs of the GP, and (3) medical students, GP-trainees, GPs, and practice nurses are made familiar with N-of-1 trials. We were unable to present an accurate estimate of the costs, because the costs of the preparation and execution of N-of-1 trials may vary according to the type of the medical problem, the people involved, and the country in which the N-of-1 trials are carried out. However, we have indicated what should be taken into account when estimating the costs (personnel and material) of an N-of-1 trial. It was also argued that health insurance companies and/or the government could bear the costs.

The implications of our results for patients are also discussed. In future research, it would be interesting to interview patients, like we interviewed GPs, and to discuss their opinions concerning the benefits, motivations, and barriers with regard to participating in N-of-1 trials. Furthermore it was discussed that N-of-1 trials may lead to shared decision making, and thus to patient empowerment, provided that the patients are actively involved in defining individualised decision rules with regard to the drug treatment.

Chapter 7 concludes with suggestions for future research, including recommendations for the preparation and execution of N-of-1 trials.

Samenvatting

Dagelijks nemen huisartsen beslissingen over de optimale behandeling voor individuele patiënten. Het is algemeen geaccepteerd dat grote gecontroleerde studies (randomised controlled trials = RCTs) doorgaans het sterkste bewijs leveren voor de effectiviteit van behandelingen. Het is echter de vraag of de resultaten van RCTs wel van toepassing zijn voor elke individuele patiënt in de huisartspraktijk.

Wanneer de huisarts of de patiënt twijfelt over de effectiviteit en/of de veiligheid van een bepaald geneesmiddel en dit objectief wil evalueren, kan N=1 onderzoek in bepaalde situaties een waardevolle aanvulling zijn op RCTs. N=1 onderzoek is onderzoek dat wordt uitgevoerd bij één patiënt. Het is een zogenaamd multiple crossover onderzoek, dat uit een aantal behandelparen bestaat. De patiënt is zijn of haar eigen controle en krijgt gedurende één periode van het behandelpaar de actieve behandeling en gedurende de andere periode van het behandelpaar de controle behandeling of een placebo ("nepgeneesmiddel"). De volgorde van de behandelingen wordt binnen elk behandelpaar door het lot bepaald. De patiënt doorloopt een aantal behandelperioden achter elkaar. De patiënt, de arts en de onderzoeker weten ten tijde van het onderzoek niet wanneer de patiënt welk geneesmiddel krijgt.

Het eerste doel van dit proefschrift was om te bestuderen of de resultaten van N=1 onderzoeken een aanvulling kunnen zijn op de resultaten van RCTs voor het nemen van beslissingen over de behandeling van individuele patiënten. Hiertoe hebben wij 3 deelstudies uitgevoerd: we hebben een samenvatting gemaakt van gepubliceerde RCTs en richtlijnen over de behandeling met geneesmiddelen van patiënten met artrose van de heup of knie *(een zogenaamde systematische review,* ***Hoofdstuk 2)***, we hebben een serie N=1 onderzoeken uitgevoerd bij patiënten met artrose van de heup of knie ***(Hoofdstuk 3)*** en een serie N=1 onderzoeken bij patiënten met slaapproblemen ***(Hoofdstuk 4)***. Beide series N=1 onderzoeken zijn in de huisartspraktijk uitgevoerd.

In onze systematische review ***(Hoofdstuk 2)*** hebben we een samenvatting gemaakt van het beschikbare bewijs over de effectiviteit van paracetamol vergeleken met die van non-steroidal anti-inflammatory drugs (NSAIDs). Tevens is de kwaliteit en de inhoud vergeleken van klinische richtlijnen met betrekking tot de behandeling van artrose van de heup of knie met geneesmiddelen. Artikelen over RCTs en klinische richtlijnen werden

op systematische wijze gezocht in bibliografische gegevensbestanden en relevante websites. Tevens werden alle referenties in de gevonden artikelen, waaronder systematische reviews, gescreend op potentieel relevante publicaties. De zoekstrategie leverde 5 RCTs en 9 richtlijnen op. De gegevens van 3 RCTs van voldoende methodologische kwaliteit konden statistisch worden gecombineerd ("gepoold"). Dit resulteerde in een klein, statistisch significant verschil tussen NSAIDs en paracetamol, waarbij NSAIDs iets effectiever was in de behandeling van pijnklachten (effectgrootte of "pooled standardised mean difference" 0,33 (95% BI 0,15 – 0,51). De gepoolde schattingen voor andere uitkomstmaten, zoals klachten of beperkingen in het dagelijks functioneren waren klein en meestal niet statistisch significant. De systematische review laat zien dat NSAIDs iets effectiever zijn dan paracetamol. NSAIDs gaan echter gepaard met een hoger risico op maag- en darmproblemen, terwijl paracetamol geassocieerd wordt met heel weinig bijwerkingen. Om deze reden raden de meeste richtlijnen voor artrose paracetamol aan als eerste keus geneesmiddel.

Aanvullend op de systematische review, waarbij de gemiddelde effecten van een behandeling onderzocht werden, hebben we twee series N=1 onderzoeken uitgevoerd om de effecten van een behandeling voor individuele patiënten te evalueren. De eerste serie staat beschreven in ***Hoofdstuk 3***. In deze serie werd voor 13 individuele patiënten die regelmatig NSAIDs gebruikten onderzocht of paracetamol even effectief was als NSAIDs bij de behandeling van pijn of functionele klachten gerelateerd aan artrose van de heup of knie. Gegevens werden verzameld door middel van dagboeken en door middel van huisbezoeken door de onderzoeker aan het eind van elke behandelperiode (iedere twee weken). De primaire uitkomstmaten waren *(a)* de ernst van de belangrijkste individuele klachten en *(b)* de intensiteit van de pijn. De secundaire uitkomstmaten waren *(a)* tevredenheid over het geneesmiddel en *(b)* een functietest (2,5 meter heen en weer lopen, 5 keer zitten en weer opstaan, of traplopen [tijd in seconden]). De patiënt voerde elke twee weken dezelfde functietest uit. Verder werd er elke twee weken informatie verzameld over het vóórkomen van bijwerkingen en over co-interventies (waaronder co-medicatie en overige behandelingen). Zes patiënten maakten het N=1 onderzoek niet af: 3 vanwege pijnklachten, 1 vanwege stijfheid van de knie, 1 vanwege ervaren bijwerkingen en 1 vanwege een andere aandoening. Voor de overige zeven patiënten liet onze studie zien dat tussen de verschillende patiënten een grote variatie bestond in de effectiviteit van

NSAIDs vergeleken met die van paracetamol: van geen tot grote verschillen in het voordeel van NSAIDs. N=1 onderzoek kan daarom gebruikt worden om te onderzoeken welke behandeling voor een bepaald persoon het beste is en daarmee kan voorkomen worden dat een patiënt onnodig lang met NSAIDs behandeld wordt. Onze studie liet echter tevens zien dat voor patiënten praktische redenen een rol kunnen spelen bij de beslissing wel of niet over te stappen van NSAIDs op paracetamol (bijvoorbeeld het aantal tabletten dat moet worden geslikt).

De tweede serie N=1 onderzoeken staat beschreven in ***Hoofdstuk 4***. Er bestaat twijfel of benzodiazepinen nog steeds effectief zijn voor slaapproblemen wanneer de patiënten deze slaapmiddelen al langere tijd gebruiken. Bovendien is het gebruik van benzodiazepinen geassocieerd met cognitieve bijwerkingen (zoals sufheid overdag), en een verhoogd risico op motorongelukken, valpartijen en fracturen. Dit is de reden waarom in richtlijnen voor de behandeling van slaapproblemen wordt geadviseerd het langdurig gebruik van bezodiazepinen te voorkómen of te stoppen. Pogingen van huisartsen en hun patiënten om na langdurig gebruik het gebruik van benzodiazepinen te stoppen zijn echter vaak niet succesvol. Het grootse obstakel is wellicht dat de patiënt weet dat hij/zij geen benzodiazepinen ingenomen heeft en daardoor minder goed slaapt. N=1 onderzoeken kunnen mogelijk meer inzicht geven in de vraag of individuele patiënten met succes het gebruik van benzodiazepinen kunnen verminderen zonder dat dit ten koste gaat van de kwaliteit en de hoeveelheid slaap. In onze serie N=1 onderzoeken hebben we voor 12 individuele patiënten die langdurig temazepam gebruikten onderzocht of temazepam even effectief was als placebo. Voor 3 andere patiënten, die het gebruik van temazepam niet wilden stoppen maar wel wilden proberen de dosering te verminderen, werd onderzocht of 10 mg temazepam even effectief was als 20 mg temazepam. Vervolgens werd onderzocht of de resultaten van het N=1 onderzoek van invloed waren op hun toekomstige gebruik van slaapmiddelen. De belangrijkste uitkomstmaten waren de individuele belangrijkste klacht en de tijd die nodig was om in slaap te vallen. Ook in deze serie N=1 onderzoeken vonden we een grote variatie in de resultaten tussen de patiënten: temazepam had bij sommige mensen geen of nauwelijks effect op de kwaliteit of de hoeveelheid slaap, terwijl bij andere deelnemers temazepam een duidelijk gunstig effect had. Onze resultaten lieten zien dat N=1 onderzoeken waardevol lijken voor patiënten die gemotiveerd zijn om het gebruik van temazepam te verminderen of ermee te

stoppen. Ze laten duidelijk de (in)effectiviteit van temazepam zien en geven patiënten mogelijk extra vertrouwen om te stoppen met het regelmatig gebruik van slaapmiddelen.

Kortom, in beide series maakten de N=1 onderzoeken het ons mogelijk de effectiviteit van geneesmiddelen voor individuele patiënten te bepalen. N=1 onderzoeken geven informatie over beslissingen rond farmacotherapie voor individuele patiënten die complementair is aan de resultaten van RCTs (waarin het gemiddelde effect voor een geselecteerde populatie wordt bestudeerd).

Het tweede doel van dit proefschrift was om de ondervonden praktische en methodologische problemen van N=1 onderzoeken te bestuderen. Vóór, tijdens en na het verzamelen van de gegevens, moesten problemen worden opgelost betreffende het meten van uitkomsten, het analyseren van de resultaten, en de uitval en de follow-up van patiënten. In ***Hoofdstuk 5*** worden onze oplossingen voor deze problemen besproken. Naar onze mening is het belangrijk de uitkomstmaten te individualiseren en aandachtig het doel van het onderzoek, de soort randomisatie en de analyse te overwegen. Tevens adviseren wij dat de patiënt het geneesmiddel in dezelfde dosering en in dezelfde toedieningsvorm krijgt als voor het onderzoek. Bovendien adviseren wij dat zowel algemene als geïndividualiseerde beslisregels worden geformuleerd die helpen bij het nemen van beslissingen ten aanzien van de effectiviteit van de behandeling, dat (zo nodig) de behandeling direct na het N=1 onderzoek wordt aangepast en dat de patiënt goede instructies en ondersteuning krijgt wanneer de behandeling wordt aangepast. Concluderend kunnen we zeggen dat N=1 onderzoeken met succes kunnen worden uitgevoerd, mits de opzet van het onderzoek goed aansluit bij de kenmerken van iedere individuele patiënt en bij het onderliggende medische probleem.

Het derde doel van dit proefschrift was om de toepasbaarheid en haalbaarheid van N=1 onderzoek in de huisartspraktijk te onderzoeken. We hebben een kwalitatief onderzoek uitgevoerd waarin een inventarisatie werd gemaakt van de door huisartsen ervaren barrières en stimulerende factoren met betrekking tot de toepasbaarheid en haalbaarheid van N=1 onderzoeken ***(Hoofdstuk 6)***. Er werd een interview gehouden met 14 van de 31 huisartsen die hadden toegestemd aan één of beide series (Hoofdstukken 3 en 4) deel te nemen. Tijdens de semi-gestructureerde interviews werd de huisartsen gevraagd het

volgende aan te geven: 1) hun positieve en negatieve ervaringen met N=1 onderzoeken, 2) motiverende en belemmerende factoren bij het uitvoeren van N=1 onderzoeken, 3) voorwaarden voor het met succes uitvoeren van N=1 onderzoeken, en 4) de eigenschappen van patiënten die geschikt lijken om aan N=1 onderzoek deel te nemen. De huisartsen van ons onderzoek beschouwden N=1 onderzoek als waardevol en nuttig voor de huisartsgeneeskunde, vanwege de kleinschaligheid en de relevantie voor de individuele patiëntenzorg. Aangezien de resultaten van N=1 onderzoek direct toepasbaar zijn voor de patiënt, vonden zij N=1 onderzoeken nuttig voor het toetsen van de behandeling bij individuele patiënten. Het werd echter ook duidelijk dat de toepasbaarheid en haalbaarheid van ieder N=1 onderzoek afhangt van een aantal factoren op het niveau van onderzoek, huisarts en patiënt. De huisartsen gaven aan dat het succes van N=1 onderzoek belemmerd kan worden door tijdgebrek, het vergeten om N=1 onderzoeken uit te voeren omdat zij niet tot de dagelijkse patiëntenzorg horen, en door de deelname van patiënten die niet gemotiveerd zijn. Ook stimulerende factoren en strategieën om deze barrières het hoofd te bieden werden geïdentificeerd. Over het algemeen vonden de huisartsen N=1 onderzoeken haalbaar en toepasbaar indien de huisarts zijn behandelplan wil laten toetsen, de N=1 onderzoeken ingebed zijn in de dagelijkse routine van de praktijk, een onderzoeksgroep goede ondersteuning biedt, de onderzoeksdoelen relevant zijn, de onderzoeksmethodologie geschikt is en alleen die patiënten meedoen die open staan voor aanpassingen in hun behandeling. Dit suggereert dat N=1 onderzoeken geïntegreerd zouden kunnen worden in de dagelijkse patiëntenzorg, met name als zij geïnitieerd en ondersteund worden door een onderzoeksgroep of N=1 service.

Tenslotte worden in ***Hoofdstuk 7*** de resultaten en conclusies van dit proefschrift kort samengevat. Vervolgens wordt aandacht besteed aan de implicaties van de resultaten van dit proefschrift voor toekomstig onderzoek. Daar onze series N=1 onderzoeken hebben laten zien dat tussen patiënten een grote variatie bestaat in de effectiviteit van geneesmiddelen, werd geadviseerd de resultaten van N=1 onderzoeken niet statistisch te combineren. Vervolgens werden farmacotherapeutische vragen en problemen besproken die met N=1 onderzoek beantwoord zouden kunnen worden. Verder werd geadviseerd RCTs uit te voeren waarin de resultaten van N=1 onderzoek worden vergeleken met die

van de gebruikelijke zorg. Het onderzoeksdoel van zo'n RCT zou zijn om te bestuderen of N=1 onderzoeken tot betere behandelingen (effectiever of veiliger) kunnen leiden en of N=1 onderzoeken kosten-effectiever zijn dan de gebruikelijke zorg .

Vervolgens wordt aandacht besteed aan de implicaties van de resultaten voor de huisartsgeneeskunde. Het implementeren van N=1 onderzoek in de huisartspraktijk kan vergemakkelijkt worden als (1) de huisartsen N=1 onderzoek beschouwen als een aanvullende test die aangevraagd kan worden (net zoals bijvoorbeeld een laboratoriumtest), (2) onderzoeksgroepen (of N=1 services) ondersteuning geven die is aangepast aan de behoeften van de huisarts, en (3) studenten geneeskunde, huisartsen (in opleiding) en praktijkassistenten bekend worden gemaakt met het N=1 onderzoek. N=1 onderzoek gaat gepaard met kosten van personeel en materiaal. In hoofdstuk 7 wordt kort aandacht besteed aan deze kosten en aan de mogelijkheden voor financiering van N=1 onderzoek.

Tot slot werden de implicaties van onze resultaten voor patiënten bediscussieerd. Voor toekomstig onderzoek zou het interessant zijn om patiënten te interviewen, zoals we ook huisartsen geïnterviewd hebben, en te vragen naar hun mening over de voordelen, de motivaties en de barrières om deel te nemen aan N=1 onderzoek. Verder werd er besproken dat N=1 onderzoek zou kunnen leiden tot " shared decision making " en daarmee tot een meer actieve rol van de patiënt bij het nemen van beslissingen rond de behandeling met geneesmiddelen.

Hoofdstuk 7 eindigt met suggesties voor toekomstig onderzoek, inclusief adviezen voor de voorbereiding en uitvoering van N=1 onderzoek.

Appendix

Setting priorities for research using N-of-1 trials in general practice: a small study to select topics for two series of N-of-1 trials

Introduction

As described in the General Introduction, N-of-1 trials may, in certain situations, be a valuable method to determine the effectiveness of a drug treatment for individual patients.[1] However, N-of-1 trials cannot be conducted for all medical problems. From a methodological point of view, a number of conditions have to be fulfilled with regard to the medical problem and the interventions. For example, the medical problem has to be chronic, and the treatment effects of the interventions should have a rapid onset and stop acting soon after discontinuation to prevent carry-over effects (i.e. the treatment effect of one treatment should have disappeared when the patient is given the other treatment). In addition, it is sensible to conduct N-of-1 trials only if there are real doubts about the treatment policy.[1] The aim of this study was to select suitable and relevant topics concerning pharmacotherapeutic doubts for a series of N-of-1 trials in general practice.

Methods

The most prevalent medical problems in the Dutch general practice were selected according to two different methods of assessment.[2,3] Subsequently, medical problems were excluded if they did not concern a complaint, an illness or a disease, if drug treatment was irrelevant, or if ethical considerations would make an N-of-1 trial unfeasible. To further prioritise the medical problems concerning their suitability for N-of-1 trials, seven GPs, who had (direct or indirect) ties with the research group and agreed to participate, were sent a questionnaire and asked to answer the following questions concerning each medical problem with " Yes ", " No ", or " Don't know ":

1) Do you frequently have doubts about the drug treatment for this medical problem for individual patients in your practice? If so, about which drugs do you have these doubts?
2) Is drug treatment for a prolonged period of time or a frequently repeated drug treatment considered?
3) Do the treatment effects of the intervention have a rapid onset (within 5 days) and stop acting soon after discontinuation?
4) Is it feasible to carry out an N-of-1 trial on this topic in the general practice setting?
5) Is it ethical to conduct the trial?

Finally, the GPs were asked to indicate the three most relevant medical problems to be studied by N-of-1 trials in general practice (Top 3).

Analysis

For each medical problem a total score was calculated, indicating the suitability of this problem for N-of-1 trials. This total score could range from 0 to 35 (5 questions x 7 positive responses). The number of GPs who placed the medical problem in the Top 3 of most relevant medical problems for N-of-1 trials was also calculated.

Results

Combining the 20 most prevalent medical problems from both assessments resulted in a total of 30 different medical problems. Subsequently, nine were excluded: three because the problem did not concern a complaint, an illness or a disease, five because drug treatment was irrelevant, and one for ethical reasons. The medical problems for which the N-of-1 trials were most suitable, according to the GPs, were: osteoarthritis, sleep disturbances, and anxiety/distress. Sleep disturbances and anxiety/distress were placed most frequently in the Top 3 most relevant medical problems (Table 1).

Table 1. Suitability and relevance of prevalent medical problems in general practice for N-of-1 trials concerning pharmacotherapeutic doubts (survey among 7 GPs)

Medical problem	Conditions for conducting N-of-1 trials (No. of GPs with positive responses)						No. of GPs who indicated the medical problem in the Top 3
	Doubts?	Prolonged/ frequent?	Effect soon?	Feasible?	Ethical?	Total score	
Osteoarthritis	5	7	7	7	7	33	1
Sleep Disturbances	6	2	7	6	7	28	4
Anxiety, distress	6	3	6	6	7	28	3
Depression	5	7	0	6	6	24	2
Functional-nervous complaints	6	2	4	4	6	22	2
Migraine	5	4	4	4	4	21	1
Low back pain without radiation	4	1	6	4	4	19	2
Diabetes Mellitus II	1	5	4	4	5	19	1
Hay fever, allergic rhinitis	2	4	4	3	5	18	0
Chronic gastritis	3	4	4	4	3	18	2
Hypertension	2	4	2	5	4	17	1
Constitutional eczema	1	5	3	4	4	17	0
Contact eczema	1	3	3	4	5	16	0
Acne vulgaris	3	5	0	4	4	16	0
Irritable bowel syndrome	5	1	1	3	5	15	2
Asthma	2	4	4	3	2	15	1
COPD	2	5	3	3	2	15	1
Angina	2	4	5	2	1	14	0
Hypercholesterolaemia	2	5	0	3	3	13	1
Psoriasis	0	4	0	2	3	9	0
Congestive heart failure	0	3	4	1	0	8	0

Doubts? = Does the GP frequently have doubts about the drug treatment for this medical problem for individual patients in the practice?; Prolonged/frequent? = Is drug treatment for a prolonged period of time or a frequently repeated drug treatment considered? Effect soon? = Do the treatment effects of the intervention have a rapid onset (within 5 days) and stop acting soon after discontinuation?; Feasible? = Is it feasible to carry out an N-of-1 trial on this topic in the general practice setting?; Ethical? = Is it ethical to conduct the trial? The higher the total score, the more suitable the medical problem for evaluation by N-of-1 trials. The more frequently placed in the Top 3, the more relevant the medical problem for evaluation by N-of-1 trials.

Conclusions

This study was carried out to select suitable and relevant topics concerning pharmacotherapeutic doubts that could be addressed by N-of-1 trials in a general practice setting. Of the most prevalent medical problems in general practice in The Netherlands, osteoarthritis was considered by the participating GPs to be the most suitable, and sleep disturbances to be the most relevant for a series of N-of-1 trials. Sleep disturbances also scored high for suitability. It was therefore decided to conduct two series of N-of-1 trials, one in patients with osteoarthritis of the hip or knee *(Chapter 3)* and one in long-term users of the hypnotic temazepam *(Chapter 4)*.

References

1. Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D, Keller J. A clinician' s guide for conducting randomized trials in individual patients. CMAJ 1988;139:497-503.
2. Lamberts H, Brouwer HJ, Mohrs J. Reason for encounter-, episode- and process-oriented standard output from Transition project. Part I. Amsterdam: Department of General Practice/Family Medicine, University of Amsterdam, 1991.
3. Van der Velden J, De Bakker DH, Claessens AAMC, Schellevis FG. The Dutch National Survey in General Practice. [In Dutch: Een nationale studie naar ziekten en verrichtingen in de huisartspraktijk.] Utrecht: NIVEL, 1991.

Dankwoord

Ja, ... het is echt af!!! Dat dit proefschrift nu voor u ligt is natuurlijk niet alleen aan mij te danken. Ik heb veel hulp gehad en wil daarom graag een aantal mensen hiervoor bedanken.

Op de eerste plaats wil ik Daniëlle van der Windt, mijn copromotor, bedanken. Daniëlle, een betere copromotor kun je je niet wensen. Naast dat je erg veel kennis in huis hebt en dit goed kan overbrengen, stond je altijd voor me klaar. Wanneer het weer eens tegen zat, hielp je me direct en wist je me telkens weer te motiveren. Bedankt!

Theo de Vries, mijn eerste promotor, vanaf het allereerste begin was je erg enthousiast over het N=1 onderzoek en dat enthousiasme heb je goed op me weten over te brengen. Je was erg betrokken bij het project en dat waardeer ik enorm (welke andere promovendus ziet zijn promotor zo vaak?). Vaak had je een andere invalshoek dan ik, waardoor onze ideeën elkaar goed aanvulden en een beter resultaat bereikt werd. Bedankt voor al je goede ideeën, betrokkenheid en stimulerende gesprekken!

Wim Stalman, mijn tweede promotor, je had altijd goede ideeën (waarom was ik daar zelf niet opgekomen?) en wist die ook helder en duidelijk over te brengen. Vooral voor het onderzoek over de toepasbaarheid van N=1 onderzoek in de huisartspraktijk heb ik vaak met je van gedachten gewisseld. Het project was niet eenvoudig, maar je hebt me op de juiste gedachten weten te brengen. Ik heb veel van je geleerd.

Marten de Haan, aan het begin van mijn promotie-traject trad je op als een interim tweede promotor. Ik wil je hartelijk danken dat je deze taak voor je rekening wilde nemen. Bedankt voor je commentaren en goede suggesties.

Caspar Chin A Fo, stagiair geneeskunde, ik wil je bedanken voor je inzet en enthousiasme. Voor de serie N=1 onderzoeken bij patiënten met artrose deed je voor een deel van de patiënten al het veldwerk en ook nam je het invoeren van de gegevens voor je rekening, bedankt!

Monique Bongers, liep voor de huisartsenopleiding stage bij de serie N=1 onderzoeken bij patiënten met slaapproblemen. Monique, je was enthousiast en erg doortastend. De deelname van een groot deel van de patiënten heb ik waarschijnlijk aan je overredingskracht te danken. Het was fijn om met je samen te werken.

Alle overige co-auteurs wil ik bedanken voor hun enorme inzet. Myrra Vernooij, ik vond het fijn om met je samen te werken en door jou weet ik nu een stuk meer over kwalitatief onderzoek. Walter Devillé en Jos Twisk, N=1 onderzoek was weer eens wat anders voor jullie. Bedankt voor jullie nuttige commentaren en het altijd voor me klaar staan. Maurits van Tulder, jouw review-ervaringen kwamen goed van pas. Bedankt voor het vele " scoren " en je goede suggesties voor de review.

Alle leden van de begeleidingscommissie wil ik graag bedanken voor hun inzet. Rob Henning, je had vaak hele duidelijke en nuttige opmerkingen; Clemens Ham, ik had veel aan je opmerkingen gebaseerd op je ervaringen uit de huisartspraktijk; Rob Scholten, naast je vrolijkheid kwamen vooral je ervaringen met reviews van pas; Walter Devillé bedankt voor het je verdiepen in de statistiek van N=1 onderzoek; Ton Schoffelmeer, bedankt voor je commentaren vanuit de farmacologie; Kees Hol, ik wil Leo Pharma bedanken voor de financiering van dit project en daarnaast wil ik je persoonlijk bedanken voor je hele praktische tips.

Natuurlijk wil ik graag alle deelnemende huisartsen en patiënten bedanken. Zonder jullie zou dit proefschrift niet bestaan.

Tevens gaat mijn dank uit naar alle medewerkers van de apotheek van het VU medisch centrum die mijn studie mogelijk hebben gemaakt. Bedankt voor jullie nuttige adviezen en voor het vullen van de medicijncassettes! Ik vond het fijn met jullie samen te werken.

Ik wil graag Patrick Onghena, Ad Sitsen, Pim Assendelft, Rob Henning, Riekie de Vet en Marten de Haan bedanken voor het beoordelen van mijn proefschrift en jullie aanwezigheid tijdens de promotie plechtigheid. Samen met Patrick Onghena heb ik een

cursus over N=1 onderzoek mogen geven in Antwerpen. Patrick, bedankt voor de fijne samenwerking en de antwoorden op mijn vragen over de randomisatie test.

Mijn kamergenoten Natasja van Schoor en Vianda Stel wil ik bedanken voor alle epidemiologische tips en voor de gezelligheid. Jullie zijn echte vriendinnen. Natasja, Vianda en onze " buren " Carry Renders en Irene van Valkengoed wil ik graag bedanken voor de gezellige etentjes. Dat moeten we snel maar weer eens doen! Met Arlette Hesselink en Petra Jellema heb ik korter op een kamer gezeten, maar ook jullie wil ik bedanken voor jullie interesse in mij en mijn project.

Valentien Blom, ik wil je bedanken voor de " kopjes thee " in de wat moeilijkere tijden, maar ook voor die in de makkelijkere tijden daarna. Ook alle overige EMGO-ers " van de gang " wil ik hartelijk danken voor de goede sfeer die mede aan het slagen van mijn proefschrift ten grondslag ligt.

Belangrijk, maar vaak onderschat zijn de secretaresses. Lieke Glaudemans (Medische Farmacologie / Farmacotherapie), Els Pronk (EMGO) en Ellen Visser (EMGO), bedankt voor jullie hulp voor van alles en nog wat (en daar valt heel veel onder).

Wim Kraan, Len Wallast, Patrick Roest en Kees de Boer wil ik graag bedanken voor het oplossen van allerlei computer-troubles tijdens mijn onderzoek.

Lex Bouter, ik wil je graag bedanken voor je interesse in mijn project. Ondanks dat ik niet een van " jouw " promovendi ben, wist je toch veel over mijn project. Tijdens de postdoctorale opleiding epidemiologie heb ik veel van je geleerd. Je bent een echte leermeester.

De SA's, oftewel de student assistenten van de afdeling Medische Farmacologie / Farmacotherapie Milan Richir, Roel Verbaak, Jelle Tichelaar, Peter van Maanen, Zana Jaf, Twana Jaf , Coen van Kan en Chris Zonnenberg wil ik graag bedanken voor hun interesse in en enthousiasme voor mijn promotie-vorderingen. Ook Joke Vollebregt en Ton Schneider wil ik graag bedanken voor hun enthousiasme. Ik zal nooit vergeten dat

Ton de laatste maanden iedere week vroeg of ik weer een artikel gepubliceerd had. Jammer genoeg, of misschien ook wel gelukkig, gaat dat niet zo snel. Ook alle andere medewerkers van de afdeling Medische Farmacologie wil ik hartelijk danken, met name voor jullie aanmoedigingen tijdens de laatste puntjes op de bekende "i ".

Nicole en Monica, jullie zijn al vriendinnen vanaf de lagere school. Jullie mannen Ronald en Richard ken ik inmiddels ook al jaren en na al die jaren weet je wie je echte vrienden(innen) zijn: Jullie dus! Heel erg bedankt voor jullie onvoorwaardelijke steun! Menno en Perrie, ik ken jullie nog maar een paar jaar, maar weet nu al dat ik op jullie kan bouwen. Bedankt voor alle gezellige kook-feestjes! Kees van de Veer, ook jou wil ik hartelijk bedanken voor je steun. Je staat altijd voor me klaar. Ik vind het dan ook erg leuk dat je mijn paranimf wilt zijn.

Dit proefschrift had ik nooit kunnen schrijven zonder mijn ouders. Pa en ma, jullie steun, liefde en vertrouwen in mij heeft het mogelijk gemaakt dat dit boekje uiteindelijk voor jullie neus ligt. Ook wil ik graag mijn broers, zussen, schoonzussen en zwager bedanken voor jullie steun en het aanhoren van mijn proefschriftperikelen. Carolien en Toos, bedankt voor de ontelbare "hoe gaat het ermee "telefoontjes. Ze gaven mij heel erg veel energie. Carolien, Ton, Marc en Monique, Toos en Marcel, Maarten en Eva, Ad en Claudia, bedankt voor jullie interesse en de nodige pep-talks. Marc, tijdens mijn "inhaalrace "op de hogere school stond je altijd al voor me klaar (zelfs heel vroeg in de ochtend om 6.00 uur) om me zonodig hierbij te helpen. Bedankt! Ik vind het dan ook erg fijn dat je mijn paranimf wilt zijn. Pa Sabajo en Helen, jullie zijn erg lieve mensen en ik voelde me meteen bij jullie thuis. Bedankt voor jullie steun.

Lieve Jan, ik wil je bedanken voor je liefde en steun. Je hebt me helpen relativeren, zodat ik met slechts weinig stress het schrijven van mijn proefschrift kon afronden. Je hebt me geleerd om juist ook NU van het leven te genieten. Onze Salsa-uitjes hebben daar zeker aan bijgedragen, maar ook onze reisjes naar verre en minder verre oorden (ik hoop dat er nog vele volgen). Bedankt, … voor alles, ik hou van je.

About the author

Anke Wegman was born in Amsterdam, the Netherlands, on December 15, 1971. In 1990, she passed at the 'Katholiek College Amsterdam-West' in Amsterdam (pre-university education). She studied medicine at the VU University Medical Center. After graduation, she worked about a year as a junior investigator at the SEO Amsterdam Economics of the University of Amsterdam. In 1998 she started as a PhD student at the Institute of Extramural Medicine (EMGO Institute) at the VU University Medical Center in Amsterdam, where she performed the research described in this thesis in co-operation with the Department of Medical Pharmacology / Pharmacotherapy at the same university. During her PhD project, she attended the statistical and methodological courses organised by the Postgraduate Epidemiology Programme of the EMGO Institute. The last 2 years, she worked two years at the Department of Medical Pharmacology / Pharmacotherapy, where she mainly co-ordinated the education programme of pharmacotherapeutics for medical students. During this period she finished this thesis.