General discussion
In this thesis, the impact of RA was assessed through five components of the International Classification of Functioning, Disability and Health (the WHO ICF) framework (2001): ‘body functions and structures’, ‘activities’, ‘participation’, ‘environmental factors’ and ‘personal factors’.\[1\] Section I (The COBRA-light trial) focused on the components ‘body functions and structures’ and ‘environmental factors’, in this case the financial impact of COBRA-light therapy. In section II (Rheumatoid Arthritis and Work), attention was paid to the components ‘activities’, ‘participation’, and ‘personal factors’.

In this chapter, the main findings will be summarised and discussed per part, as well as the limitations and comparison with literature are also presented. In the last part of this chapter implications for future research are presented.

MAIN FINDINGS AND DISCUSSION

SECTION I  THE COBRA-LIGHT STRATEGY

*Chapter 1.1 Non-inferiority of COBRA-light after 26 weeks of treatment*

In this first chapter of this thesis, the primary outcome of the COBRA-light trial, change in disease activity score of 44 joints (DAS44), at week 26 is described.[2] The results revealed that the DAS44 significantly decreased in both groups: in COBRA-light with –2.2 points (SD 1.1); in COBRA with –2.5 points (SD 1.2). The difference between the groups in DAS44 change was 0.3 point (95% CI: –0.0 to 0.7; p=0.08). All other outcome measures were not significantly different between both strategies. We could conclude that COBRA-light is non-inferior to standard COBRA therapy.

*Chapter 1.2. Non-inferiority of COBRA-light after 52 weeks of treatment*

After 52 weeks of treatment with COBRA or COBRA-light strategy, the change over time in DAS44 between baseline and week 52 was –2.0 points (SD 1.0) in the COBRA-light group, and –2.4 points (SD 1.2) in COBRA.[3] Just as after the first 26 weeks of the trial, this difference was not statistically significant. Also, 70% of the patients had no radiological progression of joint damage (with no difference between COBRA-light and COBRA strategy). Less than 5% had progression of joint damage above the smallest detectable change (increase
of the total Sharp van der Heijde score (SHS) with 5 points or more). In total, 6% of the patients had erosive disease at week 52.

Although both strategies showed good results on DAS44 scores, 75% (n=61) in the COBRA-light group, and 59% (n=47) in the COBRA group still needed intensification of their treatment with etanercept, as this percentage of patients did not reach the predefined level of DAS44 < 1.6. Remarkably, starting with etanercept was often not implemented by the treating rheumatologists, mostly because they did not agree with the assessment of DAS44 performed by the study nurses. In 40 out of the 61 patients (66%) in the COBRA-light group compared to the COBRA group 27 out of 47 patients (57%) who should have been prescribed etanercept, did actually receive the TNF blocking agent. Thirteen weeks after indication to start with etanercept, mean disease activity decreased regardless of whether etanercept was actually started. The DAS44 in the COBRA-light group decreased by 0.55 (0.8) and 0.29 (0.9) points, respectively; in the COBRA group with 0.57 (0.8) and 0.23 points (0.5) for actual starters and non-starters, respectively. A total of 46 patients received etanercept for 26 weeks (COBRA-light: n=30 and COBRA: n=16); in this group the DAS44 increased again after start of etanercept, between week 39 and 52, resulting in a net decrease of DAS44 of 0.36 points in COBRA-light and 0.31 points in COBRA therapy over the total second 6-month period.

Chapter 1.3 Cost-utility of COBRA versus COBRA-light
As chapter 1.2 showed, we concluded that COBRA-light is non-inferior to COBRA therapy on all clinical outcome measures. In line with this conclusion no difference in QALYs (based on the EQ-5D score) was found between both groups (2.5 points; 95% CI: -5.3; 10.4).

The results of the base case analysis revealed that COBRA-light strategy is more expensive than COBRA (k€9.3 (SD 0.9) compared to k€7.2 (SD 0.8)), but the difference in costs was not significant (k€2.0; 95% CI -0.3 to 4.4). This difference was mainly driven by higher direct non-medical costs after 1 year of treatment (k€1.0 (SD 1.8) for COBRA-light and k€0.3 (SD 0.6) for COBRA). Also, indirect costs were roughly 50% higher in the COBRA-light group compared to the COBRA group (non-significant): k€3.0 (SD 6.4) and k€1.6 (SD 3.9) for COBRA-light and COBRA, respectively. Analyses with other effect measures, showed similar trends.
Based on actual costs as reported by the patients, costs for etanercept were lower in the COBRA-light group, despite higher number of patients needing intensification with etanercept compared to COBRA. Due to protocol violations, less patients actually received etanercept in the COBRA-light group. Therefore, for the last sensitivity analysis, cost of etanercept use was assumed to be as indicated in the protocol. In this analysis, the costs for COBRA-light were higher compared to COBRA: €11.5 (SD 8.3) versus €8.5 (SD 6.8), respectively. This difference in costs was significant: €2.9 (95% CI 0.6 to 5.3).

With respect to the outcome measure QALY, in total 59% of the bootstrapped cost-utility pairs fell in the northwest quadrant, representing the probability that COBRA-light is more expensive and less effective compared to COBRA. For analyses with most patient outcome measures, percentage of cost-utility pairs fell mainly in the northeast quadrants (between 53% and 75%). Based on the etanercept costs, nearly all cost-utility pairs fell in the northeast quadrants for all outcome measures accept for ACR 70 response and EULAR good responder. In other words, COBRA-light is more effective, but also more expensive when compared to COBRA strategy.

**DISCUSSION**

The COBRA-light trial was a non-inferiority open-label trial. This type of trial has its limitations. Firstly, the results of open-label trials are more susceptible to bias than blinded studies. To minimize any influence on outcome assessment, these were performed by trained research nurses uninvolved in the routine care. On the other hand, the open-label design more closely mimicked daily practice, which increases its external validity. Secondly, the width of the CI is a function of power, for non-inferiority ideally set at 90% in this type of trials, whereas 80% was the maximum feasible in our setting. The relevance of non-inferiority trials depends on the choice of the non-inferiority margin.[4, 5] Based on clinical experience we arbitrarily chose a non-inferiority margin of 0.5 before starting the study. Our point estimate was well below that boundary, but the upper limit of the CI 0.53 exceeded the upper limit of 0.50 by a minimal degree. We still supported non-inferiority for the following reasons: the difference in DAS44 was largely driven by ESR; the difference between groups remained less than 0.50 at week 52; and the confidence interval did not include the 0.5.

Our results also showed that a low DAS44 threshold in treat-to-target protocols poses challenges in clinical practice, where physicians frequently
did not accept treatment intensification (with its inherent risks and costs) in the face of a satisfied patient with minimal disease activity. The measure has limited validity in patient care settings when disease activity is low: it is highly sensitive to small changes in ESR in the normal range and global health scores, when joint scores are low or zero.[6] The treatment goal in the COBRA-light trial was set at DAS44 <1.6, which is very strict compared to current practice guidelines, usually set at 2.4 - corresponding to a DAS28 of 2.6. With a combination of traditional DMARDs, 47% and 38% of patients for COBRA and COBRA-light therapy, respectively, achieved this goal compared to 32% in the BeSt trial.[7] Based on this cut-off point, 67% of our study population required etanercept treatment according to the protocol, but only 42% received it. This was also observed earlier in the DREAM study.[8] We also noticed that treating physicians sometimes did not agree with the DAS44 assessors, pointing to the well-known unreliability of joint counts, especially when these are low.

Of interest, the added benefit of etanercept appeared to be rather limited: both actual starters and non-starters ended up with a mean DAS44 score of 2.2 at week 52, suggesting anti-TNF therapy has limited benefit in patients with low disease activity. Although cost differences between both strategies were not significant, adding etanercept to the treatment resulted in an enormous increase of costs, despite limited benefit. To our knowledge, this study is the first to assess the effect of adding TNF therapy after 6 months of treatment with intensive combination regimens including prednisolone.

SECTION II  RHEUMATOID ARTHRITIS AND WORK

Chapter 2.1 Content validity of the RAID score
After performing three focus group discussions (FGDs) with 18 patients in total, we could conclude that from the seven domains in the RAID score, five were indicated as being relevant for the study population: a) coping with the disease; b) functional disability assessment (activities performed in daily life); c) pain; d) fatigue, and e) emotional well-being. The domains sleep and physical well-being were briefly or not at all mentioned in the FGDs. The domains work, relationships with third parties and leisure time activities (not in the RAID) were also considered important. Concerning comprehensibility, patients indicated having trouble interpreting the numerical rating scales as in the Netherlands.
such questions are often answered by comparing them with school grades. Patients found some questions not well formulated and in the question on physical well-being, it seems that translation errors had occurred. During the preparation of the paper, the RAID group had encountered this and the Dutch version was revised to cover the translation error.

Five out of the seven items from the RAID score refer to three domains of the World Health Organisation International Classification of Function, Disability and Health (WHO ICF) core set of RA. RAID adds two domains not covered by the WHO ICF (coping and fatigue), and omits four, of which two are outside the scope of a PRO measure (Body structures and functions).

**DISCUSSION**

FGDs are an effective method to discover people’s ideas, feelings and needs about a subject. Another advantage of FGDs is that they present a more usual setting in comparison to individual interviews as patients are influenced by and influencing others, just as in daily life, providing the opportunity for discussion and consensus.[9] It must be noted that the recruitment of patients for our study were done through an invitation letter distributed by patient organizations. It is likely that patients who have noticed that their disease altered their life situation were more likely to participate in the FGDs compared to patients who did not notice a major impact of RA on their lives. Also our patients all had established RA, perhaps limiting generalizability. We did not retrieve data of other comorbidities that patients might have. Other comorbidities might also have impact on a patient’s life.

*Chapter 2.2 Systematic review of the impact of biological therapy on sick leave*

In this chapter, the results of a systematic review, which included 19 articles (six uncontrolled cohorts, seven controlled cohorts and six RCTs), on the effect of biological treatment on work participation were described.[10] We concluded that biological therapies have an overall positive effect on sick leave and presenteeism from paid work when compared to 1) the situation before the start with a biological therapy, 2) the general population, or 3) similar groups starting of continuing usual care with disease-modifying anti-rheumatic drugs (DMARDs). However, large heterogeneity made an overall conclusion difficult.

The main problem was the use of different definitions of work participation outcome and different ways to assess these outcomes. Also, there was
heterogeneity in the study design (i.e. RCTs with limited generalizability versus cohort studies without control group).

**DISCUSSION**

This systematic review had a couple of large challenges. First, the heterogeneity in study design, work related outcomes and patients included, hampered comparability of results of all included articles. Regarding study designs RCTs typically have low external validity and follow-up is often too short to be relevant with regard to work outcome. Also, in RCTs work participation was often a secondary outcome. As a consequence, sample sizes could have been too small to show effects on employment status. Uncontrolled cohorts may be more generalizable and provide more informative with respect to the duration of the effect on work outcome, but the absence of a control group cannot exclude regression to the mean as a cause of improvement, because patients are likely to start with biological agents when the disease flares. Cohort studies that include an RA control group can adjust for the effect of non-biological treatments (e.g., non-biological DMARD); however such studies cannot adjust for societal effects on the occurrence of sick leave and employment perspectives.

Secondly, different approaches were used to assess work outcome (e.g., self-composed questionnaires, validated instruments or existing databases reporting on work outcomes). Thirdly, different concepts of absenteeism and presenteeism were presented and methods to attribute impact on work to RA or overall health differed. And finally, important determinants or confounders for work related outcomes such as educational level and job-related characteristics were never reported as baseline characteristics nor added as explanatory factors in further analyses.

**Chapter 2.3 Prediction of sick leave and worker productivity in early rheumatoid arthritis patients**

At baseline, 97 persons had a paid job, 59 patients did not have a job and 6 patients had missing information on work. The majority of the employed patients performed white collar jobs (57%), and worked ≥20 hours a week (75%). Of the unemployed patients 26 (44%) were retired, 4 (7%) had no work/ were searching for work, and 10 patients (17%) were fulltime housekeeper.

During the trial, 13 patients stopped working (8%), and 6 (4%) patients started working. The mean work days according contract did not change over
time. Mean work hours fluctuated a little, but were similar at baseline and end. Percentage of patients in sick leave decreased per time moment, and the actual days and hours decreased to a median of 0 days and hours. The worker productivity at work increased with 0.2 points on the VAS scale. Patients without a paid job or with missing data about a paid job at baseline, were excluded from further analyses.

The final prediction model for sick leave, when combining all potential personal, disease and work variables, only contained sick leave in the past three months. When this strongly predictive variable was excluded, patient global health assessment and actual hours on sick leave became predictors.

The final prediction model for improved worker productivity, when combining all potential personal, disease and work variables, contained patient global health assessment, actual hours on sick leave, a SHS score of 1 or more and higher worker productivity scores in the past three months.

**DISCUSSION**

The results of the last chapter of this thesis are not easy to generalize to other countries with different social security systems. In the Netherlands, patients become work disabled after two years of consecutive sick leave. When an employee in the Netherlands is not able to work in fulltime duty (i.e. 8 hours a day), this is defined as sick leave. Employees receive continued pay by the employer for a maximum of 2 years.[11] Usually 100% of the salary is paid for the first year and 70% is paid for the second. Secondly, the number of patients with a paid job was mainly based on a question at baseline and on information in the cost diaries. Therefore, selection bias could have occurred, as it is demanding to fill-out a diary, and requires skill.[12] Patients experiencing difficulties may not have completed any diary, or only a few. Therefore, the number of patients with a paid job could have been an underestimation of the actual patients with a paid job. Unfortunately, patients only completed information on work status at baseline, which made it impossible to pursue why patients lost their job (because of disability or retirement). Thirdly, sick leave as well as several possible predictors were dichotomized. This may lead to loss of precision, but was necessary in view of distribution of the data: e.g., many patients reported having 0 days and hours on sick leave. And finally, we collected limited information on work related characteristics. More is becoming clear on the fact that contextual factors (e.g., job control, support
from colleagues and managers) might explain the relationship with sick leave, more than disease variables.[13, 14] These factors therefore should be considered when developing strategies to prevent productivity loss, sick leave, and eventually work disability.

FUTURE RESEARCH

SECTION I THE COBRA-LIGHT TRIAL

As this thesis has shown, COBRA and COBRA-light strategies are both effective in decreasing disease activity during one year of treatment, and have similar safety profiles (chapters 1.1 and 1.2). We also found that both strategies have a favourable effect on physical activity.[15] However, data on long term effects of treating patients with high initial dosages of prednisolone are still scarce. Therefore, the COBRA-light trial was extended and patients were seen after approximately four years. Aim of this extension was to see what the effects are of both regimens, especially the high dosages of prednisolone, on body composition, bone mass and structure, cardiovascular outcome, and on safety measures. Analyses of these data are still ongoing, but when taking a glimpse on one long-term effect (bone mass), we found that patients in the COBRA-light group had more bone fractures compared to the COBRA group, despite equally results on the dual energy X-ray'-absorptiometry (DEXA) scans (preliminary data). It will be interesting to see whether we can find an explanation for the higher fracture rate in the COBRA-light group.

The COBRA-light trial applied a treat-to-target protocol (T2T). Treat-to-target recommendations were developed based on results of multiple trials, to improve treatment of RA.[16, 17] However, a large challenge in the treatment of patients in the future will be how we can enrol physicians to follow a T2T protocol, especially one based on a low DAS44 threshold. In our T2T protocol the initiation of etanercept (but also other changes in medication) was based on a DAS44 cut-off ≥ 1.6 (DAS28 ≥ 2.6) which led to many protocol violations. Based on a cut-off level of DAS44 ≥ 2.4, these number are much lower (COBRA-light: 30 (49%) and COBRA: 18 (38%)). It seems that the DAS measure is not reliable in patients with low disease activity, mostly because of its sensitivity to small changes in ESR and global health scores, when joint scores are low
or zero.(6) As reaching low disease activity is becoming highly important to show whether medication is more effective compared to existing medication, a more stable measure is preferable. In a study performed by Verhoeven et al in 2000,[18] it was shown that ACR improvement and EULAR response criteria have excellent performance, and might therefore be better indices to use for intensification or tapering of medication. It will be interesting to see whether the percentage of patients needing intensification of medication in the COBRA-light trial world decrease if the protocol had been based on either ACR improvement or on EULAR response criteria.

Many therapies are present for the treatment of RA: from the conventional DMARDs which are cheap to biological agents which are expensive. The patients in our trial had limited benefit of the addition of etanercept, which in turn led to very high direct medical costs. Therefore, a promising line of research would be to assess which patients would benefit from combination strategies including prednisolone, and who should receive a biological agent in an earlier stage. Differences between patients who responded well to COBRA-light or COBRA treatment and did not need intensification, and those who needed intensification, should be assessed on whether they differ on several outcome measures (eg., biological, genetic, and personal factors). This information could shed light on unidentified factors which could predict whether a patient will benefit from certain medication or not. If we are able to develop prediction rules which give us the chance that a patient will respond to a certain medication, treatment of patients would become personalized. This might eventually result in lower total health care costs for the treatment of RA.

SECTION II  RHEUMATOID ARTHRITIS AND WORK

The second part of this thesis focused on several patient-reported outcomes (PROs). PROs are becoming more important in the treatment of patients as they provide direct information on how the patients evaluate their health and which problems they encounter in daily life.[19-22] I have focused in this thesis on the Rheumatoid Arthritis Impact of Disease (RAID) score which was developed to measure several PROs, as well as on the impact of RA on paid work.

When focussing on the RAID score, the aim of this score was to develop an international PRO which allows us to compare patients from different
countries. Its development was based on an important construct (the impact of the disease on the patients’ life). However, its current form does not seem to cover all important factors linked to this construct, as concluded by Dutch patients with established RA (chapter 2.1). It is interesting to investigate whether patients from other countries confirm all items covered by the RAID score in its present form, or whether they also miss factors like the patients who participated in the focus group discussions. In a future update, new domains could be added, whilst other domains could be deleted or adjusted to cover the domains patients find important. This would improve the questionnaire, making it an even more promising tool to use in future trials.

One very important PRO for patients is performing paid work. Although treatment of patients has evolved in the past years and physicians are now able to better control the disease it is time to broaden our attention to the impact of the disease on social roles of the patient. The Social and Economic Council of the Netherlands is also paying more attention to the problems of performing a job with a chronic illness, and published an advisory report on managing chronic illness at work.[23] In light of this trend, the Dutch Society for Rheumatology facilitated the development of the ‘Directive for work participation and Rheumatoid Arthritis’, which was accepted in 2016 as official guideline in the treatment of RA. This directive describes how rheumatologist could pay attention to work participation during an outpatient clinic visit. A national work group (Target@Work) started an implementation project for the new directive to explain why it is important to pay attention to work, and stimulate rheumatologists to ask the patient 3-4 standard questions on work. In the centres were this project was implemented, 90% of the patients indicated to be satisfied on how their rheumatologist payed attention to work, when following the new directive. A large problem with the assessment of work as outcome, is the fact that different definitions of outcome are used. Also, work is methodologically defined differently in studies, which hampers comparison of results which also makes it difficult to come to a universal conclusion of the impact of RA on participation in work. The OMERACT (Outcome Measures in Rheumatology) group is therefore developing uniform definitions of work outcome and also investigating the best measurement instruments that can be used in trials, improving the quality of trials which asses work.

Another issue hampering comparison of trials which assessed work as outcome, is the fact that almost no study corrects for possible influential
variables such as work related factors (job characteristics) and person related characteristics (perseverance, self-esteem). Although treatment with biological treatment did show a reduction in percentage of patients on sick leave (chapter 2.2), the level of sick leave does not reach that of the general population. Insight in which factors are positively and negatively associated with work outcome will lead to better understanding of mechanisms playing a role in presenteeism, absenteeism and eventually work disability. This will allow us to develop or adapt specific intervention programs aimed at prevention of work disability, as current intervention programs often lack effectiveness, such as a recent program performed in Amsterdam by the group of Prof.dr. Adema. [24, 25] One explanation for these negative results could be that current programs might not aim at the variables which truly effect productivity at work, as they are not yet known. For that reason, future research should focus on work outcome as primary outcome, to identify possible factors which are yet unknown. Most research on work are performed as secondary or third objective. When work is the primary outcome, higher number of patients are necessary to find a strong association. Secondly, programs are now offered to every patient instead of patients at high risk for productivity loss, sick leave or work disability. Developing a prediction model which identifies patients at high risk for presenteeism or absenteeism, despite acceptable disease activity levels, would be a major step forward to help patients in remaining at work. Intervention programs could then specifically aim at these high risk patients which might lead to more effective programs. And finally, current research is performed on patients; their experiences are assessed and taking into consideration in intervention programs. Little is known about how employers and managers cope with employees who are sick (temporarily) or have chronic illnesses. It is very likely that their experiences and expectations differ from patients. This might also be a reason why intervention programs are often not effective. Knowing more of the employers’ view on chronic illness and work, might lead to intervention programs that will be more effective, resulting in less sick leave and work disability in the future.
KEY FINDINGS OF THIS THESIS

• COBRA-light is non-inferior to COBRA strategy: both effectively reduce disease activity levels, improve functional ability and slow down radiological joint progression.

• In a setting of low disease activity, etanercept has limited added benefit after 6 or 9 months of treatment with intensive combination strategies including prednisolone.

• Despite non-inferiority on clinical outcome measures, COBRA-light is more expensive compared to COBRA, mainly driven by costs due to a higher number of etanercept users and higher number of patients on sick leave.

• Sick leave in the COBRA trial participants was predicted by higher patient global assessment scores, having fatigue, higher disease duration and reporting sick leave at baseline. These factors, together with work-related factors need to be considered in strategies to prevent work disability.

• Although the RAID score offers an outcome measure solely based on patient reported outcome, patients do miss questions on work, relationships with third parties and leisure time activities.

• Biological therapies improve work participation, but the large heterogeneity of studies on this subject in terms of population, design, analyses and most importantly in outcome measures limits interpretation of the data.
REFERENCE LIST


