EVIDENCE FOR DECISION MAKING IN HEALTH CARE

BASEL INSTITUTE FOR CLINICAL EPIDEMIOLOGY AND BIOSTATISTICS

Iron therapy for iron deficiency without anaemia

Report

Publication date

April 2020

Version

2.0

Authors – Clinical effectiveness (alphabetic order)

Soheila Aghlmandi[1], Heiner C. Bucher[1], Dominik Glinz[1], Viktoria L. Gloy[1], Chandni Patel[1], Heike Raatz[1]

Head of institute

CEB: Prof. Dr. med. Heiner C. Bucher, MPH

Authors – Cost-comparison and budget impact analysis (alphabetic order)

Renato Farcher[2], Renato Mattli[2], Marco Riguzzi[2], Michael Stucki[2], Maria-Eleni Syleouni[2], Simon Wieser[2]

Head of institute

WIG: Prof. Dr. oec. Simon Wieser

[1] Basel Institute for Clinical Epidemiology and Biostatistics (CEB), University Hospital Basel and University of Basel

[2] Winterthur Institute of Health Economics (WIG), Zurich University of Applied Sciences

REAL LIFE EVIDENCE. EVIDENCE SYNTHESIS. HEALTH TECHNOLOGY ASSESSMENTS.

find out more at www.ceb-institute.org

Universitätsspital Basel

Klinische Epidemiologie

Acknowledgement

The assessment team is responsible for the content of this report. The assessment team thanks the clinical and health economic experts for their support and scientific input. The experts were recruited by the Federal Office of Public Health (FOPH/BAG) at the beginning of the project. Potential conflicts of interests have been evaluated by FOPH. The expert group has reviewed the draft scope and the draft report. Furthermore, during the scoping and assessment phases the authoring team could consult the expert group for specific questions according to their expertise. According to FOPH policy, the names of the experts are not published.

Contributions

Scoping: HR; **Systematic review:** HR wrote the protocol; DG and HR developed and conducted the literature search; DG, HR and VG screened the literature; CP and DG were responsible for data extraction; CP and DG graded the risk of bias and the quality of evidence; SA, CP and DG conducted data analysis; SA, CP, DG, HR and HCB are responsible for data interpretation; CP, DG, HR and HCB wrote the systematic review section. **Cost-comparison and budget impact analysis:** RM, MR and SW wrote the protocol; RF, MR and MES conducted the literature search; RF, MR and MES screened the literature; RM and MR summarized existing literature; RM, MR and MS built the economic model; RM, MR, MS and SW are responsible for data interpretation; RM, MR, MS and SW wrote the cost-comparison and budget impact analysis section. **Final report:** all authors agreed with the final content of the final report.

Table of contents

Abbreviations

AE	Adverse events
ADHD	Attention-Deficit Hyperactivity Disorder
CAPPS	Current and Past Psychological Survey
CI	Confidence Interval
CGI-S	Clinical Global Impression-Severity
CHF	Swiss Francs
CPRS	Conners' Parent Rating Scale
CTRS	Conner's Teacher Rating Scale
DSM	Diagnostic and Statistical Manual of Mental Disorders
DRG	Diagnosis Related Groups
EOS	End of study
FMH	Foederatio Medicorum Helveticorum (Swiss Medical Association)
FU	Follow-up
GP	General practitioner
GRADE	Grading of Recommendations Assessment, Development and Evaluation
Hb	Haemoglobin
HSR	Hypersensitive reaction
HTA	Health Technology Assessment
ICD	International Classification of Diseases
ID	Iron deficiency
IDA	Iron deficiency anaemia
IDNA	Iron deficiency no anaemia
IPD	Individual patient data
IRLS	International Restless Legs Scale
IRLSS	International Restless Legs Syndrome Severity Scale
IV	Inverse-variance
KVG	Federal Act on Health Insurance ("Krankenversicherungs Bundesgesetz")
MCID	Minimal clinically important difference
MD	Mean difference
M-H	Mantel-Haenszel methods
n	Number (of)
n.r.	Not reported
OIS	Optimal information size
par.	parenteral
PFS	Piper Fatigue Scale
PGI-1	Patient Gobal Rating of Change
PICO	Population, Intervention, Comparator and Outcomes
QoL	Quality of life
RCT	Randomised controlled trial
RLS	Restless legs syndrome
RoB	Risk of bias
RR	Relative risk ratio
SAE	Serious adverse events
SD	Standard deviation
SF	Serum ferritin
SFOPH	Swiss Federal Office of Public Health
SF-12	12-Item Short Form Survey
SMD	Standardised mean difference
SoF	Summary of Findings (GRADE output)
TSAT	Transferrin saturation
TEAE	Treatment emergent adverse events
VAS	Visual analogue scale
vs.	versus
WHO	World Health Organization

Executive summary

Background

The definition and the indication for the treatment of iron deficiency without anaemia (iron deficiency no anaemia, IDNA) are controversially discussed in Switzerland. In the past, the Swiss Federal Office of Public Health (SFOPH) has repeatedly been confronted with the question whether in specific situations iron therapy should be covered by mandatory health insurance.

Aims

In a first step the aim was to assess the clinical effectiveness of iron therapy (irrespective of the route of administration) compared to any other intervention including placebo or no therapy in IDNA populations having symptoms such as fatigue, depression, restless legs syndrome (RLS), sleep disorders, hair loss, brittle nails, attention-deficit hyperactivity disorder, and cognitive deficit. In addition to this step, an individual patient data meta-analysis of trials comparing iron therapy versus control was conducted to identify any subgroups (e.g. baseline ferritin level) in women with IDNA and fatigue who would particularly benefit from iron therapy.

In the second step, a health economic evaluation of parenteral versus oral iron therapy in symptomatic IDNA populations benefiting from iron therapy was conducted.

Methods

For the clinical effectiveness a systematic literature search was conducted in Medline and CENTRAL to identify relevant randomised controlled trials (RCTs). The systematic review was conducted according to principles of the Cochrane Handbook. Quality of Evidence was evaluated according to Grading of Recommendations Assessment, Development and Evaluation (GRADE).

For the economic evaluation, it was decided to restrict the evaluation to a cost-comparison analysis, rather than a cost-effectiveness analysis, and a budget impact analysis from a health care payer perspective because no data from RCTs with a direct comparison of parenteral and oral iron therapy could be identified and because no reliable estimation of differential effects can be expected from an indirect comparison of the available RCT data from the clinical effectiveness assessment (step one). For the cost-comparison, the medical costs of all health care services of the different routes of iron administration were modelled with a decision tree over a time horizon of one year reflecting the current clinical practice in Switzerland. The model was parametrized primarily with empirical evidence from the clinical trials identified in step one of this HTA report, from additional clinical literature and from opinions of clinical experts. The budget impact analysis was based on the results from the cost-comparison analysis, epidemiological data available for Switzerland and expert opinions.

Results

In the clinical effectiveness assessment, three symptomatic IDNA populations were identified. Eight RCTs investigated adults with restless legs syndrome (RLS), four RCTs women with fatigue and one RCT children with attention-deficit hyperactivity disorder (ADHD). In patients with RLS (eight RCTs), iron therapy compared to control let to a statistically significant reduction of RLS symptom severity and a statistically significant improvement in RLS treatment response. A potential “placebo effect” cannot

be excluded in six out of seven trials reporting on RLS symptom severity. For the outcomes sleep, sleepiness, quality of life, global impression, depression and fatigue no statistically significant effect for iron therapy compared to control was found. In women with IDNA and fatigue (four RCTs), iron therapy compared to control statistically significantly improved fatigue severity (measured as a continuous variable), improved subscores for mental and physical health quality of life and anxiety. A potential "placebo effect" cannot be excluded in the trials reporting on fatigue severity. For the outcomes fatigue improvement (measured as binary variable), quality of life total scores, and depression scores no statistically significant effect was found for iron therapy compared to control. In the individual patient data meta-analysis in 657 out of 718 (91.5%) women with IDNA and fatigue from all trials, no association between ferritin concentrations at baseline and the standardized difference of fatigue severity was found. The majority of the women (95.4%) had a baseline ferritin concentration below 50 µg/l, i.e. 74.9% of the women had a baseline ferritin concentration <30 µg/l, 20.5% between 30 and <50 µg/l and 4.6% above ≥50 µg/l. A multilevel linear regression model was used to analyse the individual patient data for fatigue severity and the model was adjusted to length of follow-up, group assignment and route of iron administration. Also, women with baseline ferritin concentration <16 µg/l had no statistically significant benefit than women with a ferritin concentration ≥16 µg/l, and women with ferritin concentration <30 µg/l had no benefit when compared to women with a ferritin concentration ≥30 µg/l. In children with ADHD (one RCT), iron therapy compared to control did not statistically significantly reduce ADHD severity or improve the clinical global impression, but statistically significantly reduced the number of children with the diagnosis of RLS. Adverse events and serious adverse events were pooled across all three study populations due to the very low numbers (only seven RCTs reported safety outcomes) and no statistically significant increase in adverse events and serious adverse events in patients treated with iron therapy compared to control was observed.

The cost-comparison analysis estimated total direct medical costs for first-line parenteral iron therapy at CHF 561 and for first-line oral at CHF 182 (time horizon one year, reference year 2018). This equals a cost difference of CHF 379 between the two treatment strategies. The univariate sensitivity analysis showed that dosage of the parenteral administration (impact +/-21.2%), duration of visit for a parenteral treatment (impact +14.8%; no lower bound defined) and probability of experiencing a severe hypersensitive reaction (impact -5.4%; +6.4%) had the largest impact on the results. In the probabilistic sensitivity analysis, the estimated cost difference between the two treatment strategies (first-line parenteral and first-line oral iron therapy) varied between CHF 304 and CHF 514 per patient in 95% of all model runs, indicating substantial uncertainty.

For the budget impact analysis, it was assumed that 24.4% instead of 0% of patients with IDNA would have been treated with first-line parenteral iron in Switzerland in 2018. This led to additional costs of CHF 10.3 million from a health care payer perspective. Considering the uncertainty regarding the size of the target population and the uncertainty in the cost difference between the two treatment strategies, these additional costs were estimated between CHF 3.3-25.0 million for the chosen time horizon. Assuming a rather hypothetical extreme scenario, meaning that all patients in 2018 would have been treated with first-line parenteral instead of first-line oral, this would have led to additional costs of CHF 42.4 million. Considering the uncertainty, these additional costs were estimated between CHF 13.6-102.6 million.

Conclusion

Although the overall quality of evidence from trials in patients with IDNA and fatigue or RLS was judged to be very low, it is likely that a substantial proportion of patients may experience a reduction in fatigue severity or RLS symptom severity from iron therapy (irrespective of the route of administration). In addition, evidence from the individual patient data meta-analysis in women with fatigue indicate that ferritin concentration at baseline is not associated with the magnitude of fatigue severity reduction.

From a health care payer perspective, the costs per patient were substantially higher for first-line parenteral compared to first-line oral iron therapy. However, the cost difference between the two treatment strategies and their budget impact were subjected to substantial uncertainty.

Preamble

The scoping report[1] by the Swiss Federal Office of Public Health (SFOPH) on iron therapy for iron deficiency without anaemia (iron deficiency no anaemia, IDNA) raised several questions. The available evidence is assessed with a multi-phased approach as described in the scope for the clinical effectiveness assessment available on the SFOPH homepage[2]. The present report covers the first phase which aims to identify high quality evidence on the effectiveness of iron therapy for symptomatic IDNA followed by an assessment of the diagnostic markers and an economic evaluation of oral versus intravenous iron therapy for those populations for which a treatment effect is being shown. For more details, consult the scope for clinical effectiveness[2] and economic evaluation[3] published on the SFOPH homepage.

Subsequent phases not covered by the present report will address following topics in more detail or based on other types of evidence: appropriateness of diagnostic and/or predictive markers and thresholds for the identification of patients who suffer from iron deficiency and are most likely to benefit from iron treatment; additional effectiveness data; evidence on the possible pathophysiology that associates iron deficiency with the conditions (with special consideration of the role of iron with regard to myoglobin and as co-factor for CNS development in children); and data on patient preferences.

1 Medical background

The definition and the indication for the treatment of iron deficiency without anaemia (IDNA) are controversially discussed in Switzerland. In the past, the Swiss Federal Office of Public Health (SFOPH) has been repeatedly confronted with the question whether in specific situations therapeutic iron therapy should be covered by mandatory health insurance. Several cases have already been submitted to courts at the cantonal level (Sozialversicherungsgerichte). According to the Federal Act on Health Insurance ("Bundesgesetz über die Krankenversicherung", KVG) a condition eligible for reimbursement has to qualify as a disease and the effectiveness, cost-effectiveness and appropriateness of its treatment must be established.

Several symptoms including fatigue, depression, RLS, sleep disorders, hair loss, brittle nails, attention-deficit hyperactivity disorder, and cognitive deficits have been put forward to be associated with iron deficiency and to represent indications for iron therapy. So far, the effectiveness of iron therapy for patients presenting with symptomatic IDNA is unclear and there has been no consensus regarding the relevant diagnostic markers and thresholds that should be used to diagnose IDNA[4-6].

The WHO defines iron deficiency as a serum ferritin concentration of <15 µg/L[7], however, it is unclear whether symptomatic populations with serum ferritin of <50µg/L would also benefit from iron therapy[4,8-10]. In order to account for this diagnostic uncertainty in this report no cut-off for serum ferritin or other blood parameters was used to quantify iron deficiency in IDNA patient populations. Therefore, any patient population without anaemia, but experiencing symptoms potentially suggestive for iron deficiency was of interest for this report.

2 Clinical effectiveness

2.1 Aim

The aim of this systematic review was to assess the effectiveness of iron therapy in patient populations having symptoms such as fatigue, depression, RLS, sleep disorders, hair loss, brittle nails, attention-deficit hyperactivity disorder, and cognitive deficit that may be suggestive for iron deficiency in the absence of anaemia.

2.2 Methods

2.2.1 Overview of the eligibility criteria

The overview of eligibility criteria (PICO-Question) used in the literature selection process is shown in Table 1.

Table 1: PICO-Question for the assessment of clinical effectiveness

Population	Adults, children and adolescents with symptomatic iron deficiency without anaemia (see section 2.2.2.1)
Intervention	Therapy with iron (see section 2.2.2.2)
Comparator	Any other intervention including placebo or no therapy (see section 2.2.2.3)
Outcomes	Health and safety outcomes (see section 2.2.2.4)
Study design	Randomised and quasi-randomised controlled trials (see section 2.2.2.5)
Languages	English, German, French, Italian (see section 2.2.2.6)

2.2.2 Eligibility criteria

2.2.2.1 Population

Studies investigating patients with symptomatic IDNA, irrespective of the definitions used for iron deficiency and the thresholds used to define anaemia, were included. Hence, no thresholds for iron deficiency or anaemia were defined for the study selection, i.e. a studies was eligible if their study population was reported to be iron deficient irrespective of the laboratory parameters for iron deficiency. Studies investigating any type of symptom were eligible. Only trials in developed countries were included. In cases where no diagnostic criteria for iron deficiency were reported, the fact that iron therapy was being investigated as a possible cure served as surrogate for the presence of iron deficiency. Similarly, in cases where no minimal haemoglobin-cut-off was reported as an inclusion criterion, it was assumed that the population was not anaemic and had a normal haemoglobin (Hb). In addition, patients were not allowed to suffer from underlying conditions known to cause symptoms that iron therapy aims to alleviate.

Studies with athletes or with patients who are known to suffer from one of the following underlying diseases were excluded:

- Congestive heart disease
- Acute renal failure, chronic kidney disease, dialysis
- Chronic liver disease
- Chronic inflammatory disease in particular – inflammatory bowel disease
- Achlorhydria, atrophic gastritis, gastric resection

- Acute or chronic infection
- Malignancies

2.2.2.2 Interventions

Studies investigating any form of iron therapy (oral and/or parenteral) were included.

2.2.2.3 Comparators

Any other intervention including placebo or no therapy. No additional criteria were defined.

2.2.2.4 Outcomes

Both health outcomes (including mortality, morbidity or quality of life) and safety outcomes, such as adverse events and serious adverse events, were assessed. Patient reported outcomes had to be relevant for patients and measured with validated instruments but surrogate outcomes were also included. In general, health outcomes rather than surrogate outcomes were deemed relevant. Relevant outcomes were identified in the included studies, i.e. after full text screening was completed. The relevant outcomes were classified according to GRADE as critical and important outcomes[11-26]. Critical outcomes would have a major impact on decision making and the quality of evidence available for these outcomes is the basis for judging the overall quality of the evidence for a clinical question. The list of assessed outcomes is summarised in the results section by patient population (see sections 2.3.2.1, 2.3.3.1 and 2.3.4.1).

2.2.2.5 Study design

Relevant study designs included randomised controlled trials (RCT) and quasi-RCTs (with assignment of treatment based on, e.g., alteration or date of birth). Although the latter methods to randomise patients are deemed inadequate, these study types were considered because it can be assumed that individuals in such studies were prospectively assigned to the intervention or the comparator[27].

2.2.2.6 Languages

Trials published in English, French, German and Italian were eligible for inclusion.

2.2.3 Literature search

The literature search comprised Medline via OvidSP and CENTRAL ("Cochrane central register of controlled trials"). Clinical experts and producers of the investigational products were given the opportunity to provide information about trials that fulfilled the inclusion criteria.

The databases were searched from inception until March 2nd 2017. The search strategy combined search terms for iron interventions with a search filter for randomised controlled trials (RCTs). Specifically, the best optimized RCT filter with regard to sensitivity and specificity, by Wong et al.[28] was used for the search in Medline, i.e. "Cochrane Highly Sensitive Search Strategy for identifying randomised trials in Medline: Sensitivity- and precision-maximizing version (2008 revision)" filter combined with the search terms "random" and "randomised" were used. Details of search strategies used can be found in Appendix 1. The search strategy was not restricted to a specific patient population because any symptomatic patient group with IDNA was considered as relevant. Conference proceedings or conference booklets were not searched; moreover, trial registries were not systematically searched because of resource constraints.

Two reviewers independently screened titles/abstracts of records found in the literature search for potentially eligible studies after removal of duplicate publications. Subsequently, two reviewers independently screened the full text articles of the potentially eligible studies in order to identify

eligible RCTs. Discrepant screening results were discussed and resolved by consensus or by third party arbitration. Protocols of included RCTs were searched for within trial registries.

2.2.4 Decision on patient-relevant outcomes to be extracted

All patient-relevant outcomes were extracted and included in the assessment.

2.2.5 Data extraction

Data on study characteristics and patient-relevant outcomes (health outcomes) were extracted into a standardised form by one reviewer and checked by another. Discrepancies were resolved by discussion or third party arbitration.

Information on patient recruitment time, maximum follow-up time, setting and country, age, sex, eligibility criteria, and description of the study interventions were extracted.

Outcome data were extracted for the latest follow-up time-point. However, earlier time-points were extracted if a specific outcome was only reported at an earlier time-point, or if drop-out rates for the later follow-up time-point were high. Inclusion of these outcomes was decided on a case-by-case basis.

Continuous outcome data were extracted as mean values for each intervention group at follow-up or, if not reported, as mean change from baseline.

Adverse events and serious adverse events were extracted for safety outcomes. Therefore, the number of patients experiencing an (serious) adverse event was analysed and not the number of events themselves. If only the number of events was reported, this information was extracted and was summarized in the relevant sections, but was not used for the pooled analysis. Similar, if side effects, complications, treatment-related adverse events, etc. were reported instead of adverse event, those information were not used for the pooled analysis, but were summarized in the text.

2.2.6 Risk of bias and quality of evidence assessment

One reviewer assessed the internal validity (risk of bias assessment) of each trial. This was checked by a second reviewer. Discrepancies were resolved by discussion or third party arbitration.

To assess the risk of bias of individual trials the following criteria were used[11-27]:

- adequate random sequence generation (selection bias)
- adequate concealment of treatment allocation (selection bias)
- adequate blinding of patients and health carers (performance bias)
- adequate blinding of outcome assessors (detection bias)
- complete outcome data (attrition bias)
- reporting bias

Risk of bias for each of the aforementioned criteria was assessed as low, high or unclear in each trial. It was taken into consideration that blinding of outcome assessors is of less relevance for some outcomes (e.g. SAE) than for patient-reported outcomes. To judge the completeness of outcome data and the resulting risk of attrition bias, the following operationalisation was used:

- The risk of attrition bias was judged low if the proportion of patients with missing data was 0 - 10% in either study arm and comparable between the randomised treatment arms.

- The risk of attrition bias was also judged low if the proportion of patients with missing data was between 10-20% per arm, was comparable between the randomised treatment arms, and was being addressed using adequate methods. In case of continuous data, methods considered to be adequate were multiple imputation methods but not simple replacement methods like “last observation carried forward” or “baseline carried forward”. In case of binary data adequate methods to address missing data were conservative assumptions about missing data; i.e. those patients with missing data in the control arm are treated in the analysis as if they have had beneficial outcome results.
- Missing data in the treatment arms were considered comparable if the difference between the intervention and control group was 5% or less.
- The risk of attrition bias was judged high if more than 20% of the data were missing irrespective of how the missing data were addressed in the analysis.

Reporting bias was judged to be low in a trial if all outcomes relevant for the review were stated in both the methods section and the results section.

The quality of the evidence was judged by one reviewer and checked by another according to GRADE (Grading of Recommendations Assessment, Development and Evaluation) on the outcome level by considering all the available trials for the respective outcome. Discrepancies were resolved by consensus or third party arbitration. The following criteria were considered to judge the quality of the evidence[11-26]:

Criteria for rating down the quality of evidence:

- risk of bias (internal validity)
- inconsistency
- indirectness
- imprecision
- publication bias

Criteria for rating up the quality of evidence:

- large magnitude of effect
- dose-response gradient
- all plausible confounders or other biases increase the confidence in the estimated effect

Imprecision referred to the confidence in the effect estimate. For continuous outcomes, the precision was adequate if the optimal information size (OIS) was sufficient (simple sample size calculation to estimate whether the total number of included patients would be sufficient for an adequately powered RCT) and for binary outcomes, if the number of events was sufficient (rule of thumb >300 events)[16]. If the sample size or number of events was sufficiently large, the 95% CI of the effect estimate was examined. If the 95% CI was narrow enough not to include both the “no effect” line and a possible clinically relevant effect (also called minimal clinically important difference) precision was adequate[16].

Using the GRADEpro GDT software[29] results of the judgement were presented in a summary of findings table.

2.2.7 Data synthesis

Study characteristics and results of the eligible trials were presented per study in tables and descriptively summarised.

The main focus of the analysis was on the latest time-point that was reported; earlier time-points were included in analysis if a specific outcome was only reported to an earlier time-point, or if the later follow-up time-point had high drop-out rates.

Where possible, outcome results were summarised quantitatively in a meta-analysis by using a random-effects model. Given that the pooled trials vary in study characteristics, e.g. setting, therapy, a random-effects model which includes the assumption that the different studies are estimating different, yet related (according to a random distribution) intervention effects was chosen. In this model the inverse-variance (IV) method[30] for continuous outcomes and the Mantel-Haenszel method[27] (M-H) for binary outcomes was applied. In the IV method, the weight given to each study is chosen to be the inverse of the variance of the effect estimate (i.e. one over the square of its standard error). Thus, larger studies, which have smaller standard errors, are given more weight than smaller studies. For dichotomous outcomes the M-H method was chosen for its better statistical properties if there are only few events. The analyses were performed using Review Manager (Version 5.3.5).

If outcomes were mentioned in the included publications but relevant data were missing, study authors were not contacted. If missing standard deviations (SDs) could not be calculated based on other information given in the publication and were not provided by study authors, missing standard deviations were approximated by the median standard deviations of other included RCTs on the same outcome measure[27]. For data where it was unclear whether the mean or the median had been given, it was assumed that the data referred to the mean. For publications reporting medians in a normally distributed study population, standard deviations were calculated based on the interquartile ranges[27]. If that was not possible, other SDs reported in the publication were discussed for approximation and this was indicated in the analysis.

Continuous outcomes were presented as mean differences. For binary outcomes, relative risks were determined. Effect estimates (summary and single for each trial) with the corresponding 95% confidence interval were presented in forest plots.

If a continuous outcome was measured on different scales, mean differences of the individual trial results were standardised using the following formula:

Standardised mean difference (SMD) = $(\text{mean}_{\text{iron}} - \text{mean}_{\text{comparator}})/\text{SD}_{\text{pooled}}$

An effect size above 0.2 SDs was considered to correspond to a small effect; effect sizes above 0.5 SDs to a medium effect and above 0.8 SDs were considered to correspond to large effects[31,32].

Heterogeneity of pooled effect estimates was estimated using I^2. Estimates of I^2 were interpreted under the guidance of the Cochrane Handbook[27]. Heterogeneity with an I^2 of 0% to 40% was considered low, 41% to 60% was considered moderate, and 61% to 100% was considered high. The interpretation of the observed I^2 value depended on other measures for heterogeneity, namely Tau^2 (a Tau^2 value of 0.04, 0.09, and 0.16 represent low, moderate and high heterogeneity, respectively), the precision of the individual effect estimates of the included RCTs, and visual examination[27,33].

In case of substantial or considerable heterogeneity, methodological and clinical factors that might explain the heterogeneity were explored in subgroup and sensitivity analyses.

2.2.8 Subgroup analyses

2.2.8.1 Subgroup analyses – trial-specific (aggregated data) meta-analysis

To assess possible variations of treatment effects by the type of intervention and study design subgroup analyses were conducted for the pre-specified subsets listed below. These subgroups were also the pre-specified criteria for the exploration of heterogeneity for pooled effect sizes. In addition to these subgroup meta-analyses based on aggregated data, an individual patient data meta-analysis was considered if several criteria were fulfilled (see section 2.2.8.2).

The sequence of the subgroup analyses listed below corresponded to the sequence in which the subgroup analyses were performed depending on the available evidence.

1. Oral vs. intravenous therapy with iron (vs. intra-muscular therapy with iron)
2. Female vs. male participants
3. Ferritin concentrations, i.e. <16 vs. ≥16 and <30 vs. ≥30 and <50 vs. ≥50 µg/l
4. Adolescents/children vs. adults

Subgroup differences were assessed by interaction tests available within Review Manager 5.3 and according to the Cochrane Handbook[27].

2.2.8.2 Subgroup analyses – individual patient data meta-analysis

IPD meta-analyses were considered after the systematic search was conducted and after preliminary data from the clinical effectiveness were available (see following section 2.3). According to the scope[2] biomarkers for iron deficiency were to be evaluated in an individual patient data meta-analysis (IPD) for those patient populations and critical outcomes where a treatment effect from iron therapy was observed, in order to identify patient subpopulations that would most benefit from iron therapy. The feasibility to conduct an IPD meta-analysis was investigated by taking into account the accessibility to the individual patient trial data, the consistency of reported outcomes in the studies and the importance of iron therapy for these patient populations for the Swiss setting. After preliminary effectiveness data from aggregated data from the present report were available, the assessment team and the SPOPH decided that an IPD meta-analysis should be conducted for women with fatigue. For all trials in women with fatigue the sponsors or principal investigators were based in Switzerland. These principal investigators or sponsors were contacted. Case report forms were requested to further explore the feasibility of the IPD meta-analysis and to compose a data analysis plan. Importantly, the IPD meta-analysis was supposed to assess the association between iron deficiency biomarkers and outcomes across all trials, therefore only outcomes and biomarkers which were consistently reported by all trials were considered (see also sections 2.2.8.2.2 and 2.2.8.2.3). Based on information from the individual trial publications and case report forms, it was decided to limit the set of predictors to baseline parameters like serum ferritin, transferrin, transferrin saturation, soluble transferrin receptor, haemoglobine and erythrocyte indices of anaemia (presence of microcytosis or hypochromia). In addition, the assessment of the association between clinical patient parameters and outcomes, such as prior depression, anxiety, etc., were considered.

2.2.8.2.1 Aim of the individual patient data meta-analysis

The aim of this IPD meta-analysis was

1. to identify possible associations between levels of iron deficiency biomarkers at baseline and the decrease on fatigue severity due to iron therapy.
2. to identify patient subgroups with different levels of iron deficiency biomarkers and to look at the interaction of these markers with iron-therapy and the critical outcome.

2.2.8.2.2 Critical outcomes to be assessed with individual patient data meta-analysis

Fatigue severity was the only critical outcome that could be evaluated across all trials because other patient-relevant outcomes were not measured or inconsistently reported by the individual trials.

2.2.8.2.3 Potential predictors of treatment response

As specified above (see section 2.2.8.2) only predictors that were consistently measured in all trials were considered for the IPD analysis in women with fatigue. After receipt and inspection of the final IPD, it became apparent that no clinical patient parameters (like prior depression, anxiety, etc.), but serum ferritin, haemoglobin and the erythrocyte indices had been uniformly measured in all trials. Therefore only these uniformly measured variables could be used for IPD analysis.

2.2.8.2.4 Original trial data sets

After legal and ethical aspects had been clarified, and formal requests (e.g. data transfer agreement) with sponsors or investigators had been resolved, the fully anonymized unedited databases containing the IPD were transferred to a secure server of the University Hospital Basel with limited access. A sanity check was done to assure data completeness and queries were resolved directly with the investigators of the trials. The number of missing observations for each baseline variable and the critical outcome variable was assessed for baseline and for the same follow-up time point as in the aggregated meta-analysis (last follow-up time point). The pattern of missingness was investigated by cross-tabulating baseline variables across all trials to explore rates of missing data and whether some were systematically missing[34]. Use of data imputation technique was foreseen, however, for the IPD analysis in women with fatigue, the recording of the four data-sets from the individual trial differed substantially and did not allow to impute data without introducing further uncertainty. Therefore, the analyses were based on complete cases. Relevant variables of the individual datasets were then formatted and merged in the IPD master-file.

2.2.8.2.5 Data analysis – IPD meta-analysis

All analyses followed the intention-to-treat principle, with all patients analysed in the arm to which they had been randomised. The follow-up time points used for the analysis were the same as for the aggregated data meta-analysis. The only outcome available by all four trials (fatigue severity) was reported on different scales and therefore the outcome was standardised; i.e. individual outcome scores were subtracted by the trial mean outcome score and then divided by its standard deviation[35]. A one-step approach (with no reproduction of the individual trial results) was chosen. After visual inspection of scatter plots that presented the mean change scores in fatigue severity by baseline serum ferritin level and by trial and treatment groups, a multilevel mixed linear regression model was used with random-effects (trial level) to account for within and between trial variability. The model was adjusted to group allocation, ferritin concentration at baseline, follow-up period (in days from baseline date to follow-up date) and route of administration. Because only four trials were included, also a

linear regression model with patient-level variables using robust standard error was used to check for trial clustering effects.

The baseline variables of interest (iron deficiency biomarker) were included as continuous variables and, in an additional analysis, as categorical variables (e.g. in form of tertiles or quartiles, or for ferritin concentration, the same cut-off as predefined in section 2.2.8.1 were applied). To acknowledge falsely increased serum ferritin concentrations in patient with an ongoing inflammation (e.g. infection) at baseline or follow-up, a sensitivity analysis was considered if inflammation markers like C-reactive protein measures were available for the corresponding time points. However, C-reactive protein was measured only in three trials included in this IPD meta-analysis and in these three trials, all women with elevated CRP had been excluded. The fourth trial did not measure C-reactive protein.

IPD based sensitivity analyses included analysis by the route of iron administration and a per-protocol patient population analysis (by exclusion of protocol violators).

All analyses were performed using Stata version 15.0 (College Station, Texas) and graphical inspections were performed in R (Version, 3.4.1).

2.2.9 Sensitivity analyses – trial-specific (aggregated data) meta-analysis

In case of substantial or considerable heterogeneity, measured with I^2, and if too few RCTs were available for subgroup analysis, explorative sensitivity analyses were conducted. Sensitivity analyses might explain how specific parameters (e.g. study characteristics) might affect heterogeneity. Therefore, RCTs with characteristics (different comparator, different inclusion criteria, etc.) that varied from the other RCTs were excluded from the analysis. Sensitivity analyses were defined posterior.

2.3 Results

2.3.1 Literature search

The electronic literature search yielded 10'959 records (Figure 1). After removing duplicates, 7'350 records were screened at title and abstract level and 246 potentially relevant records were screened in full text. After full text screening, 13 RCTs (12 full research articles and five abstracts) were included. For one identified RCT, only a conference abstract was found in the original search, the full text research article was published on June 23, 2017 after the original search was completed and then included in the present report. Finally, 18 references (13 full research articles plus 5 abstracts) for a total of 13 RCTs were included (Figure 1). Details regarding the search strategy and the number of studies and publications included are documented in Appendix 1. The study selection process is presented in Figure 1.

Three patient populations with symptoms related to IDNA were identified in these trials. Eight RCTs included adults with RLS, four RCTs included women with fatigue and one RCT included children with ADHD and IDNA. No published RCTs were identified for the other pre-defined conditions (depression, sleep disorders, hair loss, brittle nails and cognitive deficits). As multiple publications were identified for some of the RCTs, a unique RCT name was assigned to each RCT throughout the report. The overview of included RCTs can be found in Table 2, Table 7 and Table 16.

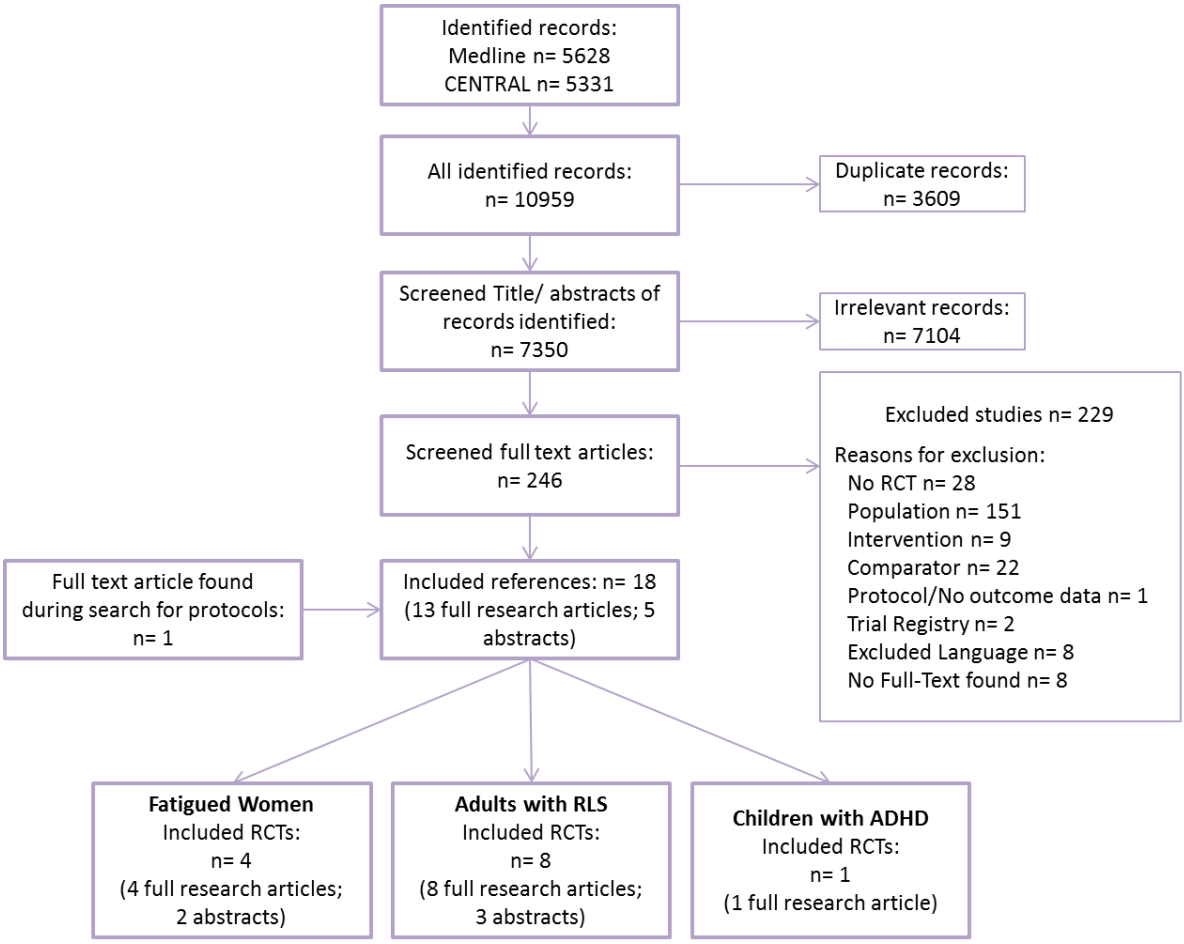


Figure 1 Trial selection process

2.3.2 Adults with restless legs syndrome

First the RCT characteristics and risk of bias assessment, and then the results for each outcome for adults with RLS are shown in the following sections. The diagnosis of RLS across all RCTs was based on the same features: urge to move legs, usually accompanied or caused by uncomfortable and unpleasant sensations in the legs; a) beginning or worsening during periods of rest or inactivity such as lying or sitting; b) partially or totally relieved by movement, such as walking or stretching, at least as long as the activity continues; and c) worsening in the evening or night compared to the day or only occurring in the evening or night[36].

2.3.2.1 Overview of included RCTs

Eleven references (eight full research articles and three abstracts) encompassing eight relevant RCTs have been identified. References can be found in Table 2. An overview of the included outcomes with analysed follow-up time-points from each RCT is given in Table 3**.**

Table 2 Adults with restless legs syndrome: Overview of included RCTs, their trial names and references

Trial name	Reference (Main reference highlighted in colour)
Allen 2011[37]	Allen RP, Adler CH, Du W, Butcher A, Bregman DB, Earley CJ. Clinical efficacy and safety of IV ferric carboxymaltose (FCM) treatment of RLS: a multi-centred, placebo-controlled preliminary clinical trial. *Sleep Med.* 2011;12(9):906-913.
Cho 2016[38,39]	Cho YW, Allen RP, Earley CJ. Clinical efficacy of ferric carboxymaltose treatment in patients with restless legs syndrome. *Sleep Med.* 2016;25:16-23.
	Cho Y, Allen RP, Earley CJ. Clinical efficacy of ferric carboxymaltose treatment in patient with restless legs syndrome. *Sleep. Conference: 30th annual meeting of the associated professional sleep societies, LLC, SLEEP 2016. Denver, CO united states. Conference start: 20160611. Conference end: 20160615. Conference publication: (var.pagings).* 2016;39:A227-a228
Davis 2000[40]	Davis BJ, Rajput A, Rajput ML, Aul EA, Eichhorn GR. A randomized, double-blind placebo-controlled trial of iron in restless legs syndrome. *Eur Neurol.* 2000;43(2):70-75.
Earley 2009[41]	Earley CJ, Horska A, Mohamed MA, Barker PB, Beard JL, Allen RP. A randomized, double-blind, placebo-controlled trial of intravenous iron sucrose in restless legs syndrome. *Sleep Med.* 2009;10(2):206-211.
Grote 2009[42]	Grote L, Leissner L, Hedner J, Ulfberg J. A randomized, double-blind, placebo controlled, multi-center study of intravenous iron sucrose and placebo in the treatment of restless legs syndrome. *Mov Disord.* 2009;24(10):1445-1452.
Lee 2014[43,44]	Lee CS, Lee SD, Kang SH, Park HY, Yoon IY. Comparison of the efficacies of oral iron and pramipexole for the treatment of restless legs syndrome patients with low serum ferritin. *Eur J Neurol.* 2014;21(2):260-266.
	Yoon I, Lee C, Lee S, Kang S, Park H. Comparison of efficacy between oral iron and dopamine agonists in the treatment of patients with restless legs syndrome with low-normal serum ferritin. *Sleep.* 2013;36:A247.
Trenkwalder 2017[45,46]	Trenkwalder C, Winkelmann J, Oertel W, Virgin G, Roubert B, Mezzacasa A. Ferric carboxymaltose in patients with restless legs syndrome and nonanemic iron deficiency: A randomized trial. *Mov Disord.* 2017.

	Trenkwalder C, Winkelmann J, Oertel W, Virgin G, Roubert B, Mezzacasa A. Single-dose ferric carboxymaltose for the treatment of restless legs syndrome in iron deficient non-anaemic patients-a randomized, placebo-controlled trial. *Journal of Sleep Research. Conference: 23rd Congress of the European Sleep Research Society, ESRS 2016. Italy. Conference Start: 20160913. Conference End: 20160916.* 2016;25:67-68.
Wang 2009[47]	Wang J, O'Reilly B, Venkataraman R, Mysliwiec V, Mysliwiec A. Efficacy of oral iron in patients with restless legs syndrome and a low-normal ferritin: A randomized, double-blind, placebo-controlled study. *Sleep Med.* 2009;10(9):973-975.

Table 3 Adults with restless legs syndrome: Overview of the included outcomes with analysed follow-up time-points

Outcome / Trial name	RLS Symptom Severity	RLS Treatment Response	Sleep	Sleepiness	Quality of Life	Global Impression Rating	Change in Global Impression, Inventory or Rating	Depression	Fatigue	Adverse Events	Serious Adverse Events
Allen 2011	4	4	4		4		4		4	4	EOS
Cho 2016	6	6	6		6						6
Davis 2000			12		12					[14]	
Earley 2009	2	2				2				[2*]	
Grote 2009	11	11								52	
Lee 2014	12	12	12	12				12			
Trenkwalder 2017	12	12		[12]		12	12			12	12
Wang 2009	12				[12]						12

The numbers in the fields denote the analysed follow-up period in weeks. Reported outcomes that could not be pooled are presented in brackets "[ ]".
*Earley 2009 reported side effects at day of infusion and adverse effects at 2 weeks.
Abbreviations: EOS, end of study; RLS, restless legs syndrome

2.3.2.2 Characteristics of the included RCTs

General characteristics of RCTs on RLS are summarised in Table 4 and selected baseline characteristics of patients from each RCT are presented in Table 5. Of the eight RCTs identified, four were from the USA, two from Korea and one from Sweden. The remaining RCT was conducted in Finland, Germany, and Switzerland. Four of the eight RCTs were conducted at single sites, while three were multicentre RCTs. In the remaining RCT, the number of RCT sites was not reported. Length of follow-up time-points extracted and analysed ranged from two weeks to 12 weeks. One RCT (Earley 2009) planned a follow-up of two years, but the RCT was stopped after two weeks because of lack of clinical benefit. Seven of the eight trials used placebo as the comparator, while one (Lee 2014) used pramipexole, a dopamine agonist used in RLS treatment, as the comparator. The 373 participants had RLS and, in all but one RCT (Davis 2000), patients were not undergoing any RLS treatment. In two of the eight RCTs (Grote 2009 and Lee 2014), only iron-deficient patients (serum ferritin ≤30/45 µg/l and 15-50 µg/l, respectively) were recruited. In two other RCTs, recruited patients had low to normal serum ferritin concentrations (Trenkwalder 2017 and Wang 2009), although only a description of iron deficiency in the recruited

population was reported in Trenkwalder 2017. In the remaining four RCTs, the iron status was considered unclear because iron status was not explicitely mentioned as inclusion or exclusion criterion. However, baseline characteristics of the patients in these four RCTs showed that Allen 2011 included patients with low ferritin concentrations, whereas in the other three RCTs the mean serum ferritin concentrations at baseline were rather close-to-normal (serum ferritin > 50 µg/l). These studies were still considered for the analysis because iron therapy was enough justification to consider the population to be iron-deficient (see also Eligibility criteria - Population 2.2.2.1). In four of the eight RCTs, a minimum cut-off for Hb-concentration was not an inclusion criterion; therefore, the anaemia status of these populations was not clear.

Four RCTs were industry sponsored (Allen 2011, Earley 2009, Grote 2009, Trenkwaldder 2017), two RCTs were supported by public funding institutions (Lee 2013, Davis 2000) and funding was unclear for two RCTs (Cho 2016, Wang 2009).

Table 4 Restless Leg Syndrome: Study characteristics

Study Name **Country**	**Setting** **Enrollment period** **Time-points of FU**	**Population** **Key inclusion criteria***	**Intervention** Compound Dosage regimen	**Comparator** Compound Dosage regimen
Allen 2011 USA	Multicentre n.r. 5 days, 2 and 4 weeks	Patients with moderate RLS IRLS score: ≥15 Fulfill NIH criteria for RLS** ID: was not an inclusion criteria Hb: was not an inclusion criteria	Ferric caboxymaltose Intravenous 500 mg in 100 ml of normal saline solution at day 0 and day 5	Placebo Intravenous 100 ml of normal saline
Cho 2016 Korea	n.r. Sept. 2013 – Oct. 2015***** (study period) 4 and 6 weeks	Patients with moderate to severe RLS Korean Hopkins–Hening Telephone Diagnostic questionnaire IRLSS scale: ≥15 RLS symptoms occurring >5 nights/week ID: not an inclusion criteria Hb ≥12 g/dl****	Ferric caboxymaltose Intravenous 1000 mg in 100 ml of normal saline solution at day 0	Placebo Intravenous 100 ml of normal saline
Davis 2000 USA	Neurology Clinic (1 site) n.r. 2, 12, 14 and 24 weeks	RLS patients without anaemia Symptomatic RLS + undergoing RLS treatment ID: was not an inclusion criteria Hb ≥10 g/dl	Ferrous sulfate Oral 325 mg, solution, twice daily	Placebo (containing 2% carboxy-methylcellulose) Oral, solution, twice daily

Study Name Country	Setting Enrollment period Time-points of FU	Population Key inclusion criteria*	Intervention Compound Dosage regimen	Comparator Compound Dosage regimen
Earley 2009 USA	General Clinical Research Center (1 site) n.r. 2 and 4 weeks, monthly until 2 years after initial treatment	Patients with RLS without anaemia Johns Hopkins telephone diagnosis interview PLMS: >15/h on second-night polysomnogram IRLS score: n.r. ID: was not an inclusion criteria Hb ≥12 g/dl	Iron sucrose (Venofer®) Intravenous 500 mg in 500 mL solution on day 3 and day 4	Placebo Intravenous 500 ml saline solution on day 3 and day 4
Grote 2009 Sweden	Multicentre (3 sites) n.r. 3, 7 and 11 weeks, then 5, 8 and 12 months	Patients with RLS IRLS score: ≥10 Fulfill NIH criteria for RLS** ID: SF ≤30/45 µg/l*** Hb: was not an inclusion criteria	Iron sucrose (Venofer®) Intravenous 200 mg at five occasions evenly spread over a 3-week period (1000 mg in total)	Placebo (sodium chloride 0.9%) Intravenous at five occasions evenly spread over a 3-week period
Lee 2014 Korea	Sleep clinic (1 site) Nov. 2010 – Jul. 2012 2, 4, 8 and 12 weeks	Patients with RLS with low-normal serum ferritin Fulfill NIH criteria for RLS** ID: SF ranging 15 – 50 ng/ml Hb: was not an inclusion criteria	Ferrous sulfate Oral 325 mg twice daily, probably entire study duration: 12 weeks	Pramipexole Starting with 0.25 mg daily, then doses titrated at every visit based on effectiveness and tolerability
Trenkwalder 2017 CH, DE, FI	Multicentre (13 sites) Apr. 2014 – Sept. 2015***** 4 and 12 weeks	Patients with moderate to severe RLS, with ID and without anaemia IRLS score: ≥15 ID: SF <75 µg/l or [≥75 µg/l and ≤300 µg/l with TSAT <20%] Hb: ≥11.5 g/dl (females) and ≥12.5 g/dl (males)	Ferric caboxymaltose Intravenous, single dose 1000 mg on day 1	Placebo (sodium chloride 0.9%) Intravenous 250 ml saline solution on day 1

Study Name Country	Setting Enrollment period Time-points of FU	Population Key inclusion criteria*	Intervention Compound Dosage regimen	Comparator Compound Dosage regimen
Wang 2009 USA	Army medical center (1 site) n.r. 6 and 12 weeks	Patients with RLS with low-normal ferritin without anaemia IRLS score: ≥11 Fulfill NIH criteria for RLS** ID: Ferritin ranging 15 – 75 ng/ml Hb: ≥11.1 g/dl (females) and ≥14 g/dl (males)	Ferrous sulfate Oral 325 mg, capsules, twice daily	Placebo (appearance-matched) Oral, capsules (Lactose)

* see Appendix 5.2 for more details on inclusion and exclusion criteria; ** see also Allen et al., Sleep Medicine, 2003; *** the initial cut-off was <30 µg/L and was increased to 45 µg/L after recruitment of 30 patients; ****In Cho 2016, an exclusion criterion for serum haemoglobin concentration of <12 µg/dl was reported; however, reviewers came to the conclusion that this was a typographical based on the author's statement of a non-anaemic population error. Therefore, the exclusion criterion for serum haemoglobin was changed from <12 µg/dl to <12 g/dl; *****In Cho 2016 and Trenkwalder 2017 only the study period was reported.
Abbreviations: CH, Switzerland; DE, Germany; FI, Finland; FU, Follow-up; Hb, haemoglobin concentration; IRLS (by Allen 2011, Grote 2009), International Restless Legs Study Group Rating Scale; IRLS (by Wang 2009), International Restless Legs Scale; IRLSS (by Cho 2016), International Restless Legs Syndrome Severity Scale; n.r., not reported; RLS, restless legs syndrome; USA, United States of America

Table 5 Restless legs syndrome: Baseline characteristics

Study Name	Intervention Group*	Comparator Group*
Allen 2011	24 randomised IRLS score: 25.1 ± 5.8 7 males (29.2%) Age: 49.5 ± 11.4 years Serum ferritin: 70.0 ± 22.8 µg/l (Male); 28.1 ± 22.9 µg/l (Female) Hb: n.r.	22 randomised/ baseline characteristics only reported for 19 IRLS score: 24.2 ± 5.5 9 males (47.4%) Age: 54.8 ± 13.6 years Serum ferritin: 58.7 ± 33.1 µg/l (Male); 24.8 ± 20.2 µg/l (Female) Hb: n.r.
Cho 2016	32 randomised IRLSS score: 27.4 ± 4.03 6 males (18.8%) Age: 49.7 ± 13.7 years Serum ferritin: 53.5 ± 41.8 ng/ml** Hb: 13.3 ± 1.42 g/dl	32 randomised IRLSS score: 28.0 ± 5.16 8 males (25.0%) Age: 52.3 ± 10.7 years Serum ferritin: 69.3 ± 55.4 ng/ml** Hb: 13.5 ± 1.11 g/dl
Davis 2000	14 randomised IRLS(S) score: n.r. 5 males (35.7%) Age: 58.6 years (33 – 80)*** Ferritin: 134.8 ng/ml (9 – 680)*** Hb: 14.3 g/dl (12.7 – 16.9)***	14 randomised IRLS(S) score: n.r. 4 males (28.6%) Age: 59.9 years (33 – 76)*** Ferritin: 100.6 ng/ml (8 – 335)** * Hb: 13.7 g/dl (11.6 – 15.6)***
Earley 2009****	n randomised: n.r. (11 received treatment) IRLSS scale: 30.8 ± 9.2 5 males (45.5%) Age: 66.4 ± 11.4 years Serum ferritin: 78.3 ± 41.7 ng/ml Hb: 15.0 ± 1.2 g/dl	n randomised: n.r. (7 received placebo) IRLSS scale: 29.7 ± 2.9 2 males (28.6%) Age: 61.4 ± 10.0 years Serum ferritin: 70.3 ± 21.5 ng/ml Hb: 14.0 ± 0.84 g/dl

Study Name	Intervention Group*	Comparator Group*
Grote 2009	29 randomised	31 randomised
	IRLS score: 24 (10-37)***	IRLS score: 26 (13-36)***
	4 males (13.8%) Age: 47 ± 10 years Serum ferritin: 20.1 ± 12 ng/ml Hb: 12.9± 1.8 g/dl	3 males (9.7%) Age: 46 ± 8 years Serum ferritin: 20.4 ± 11 ng/ml Hb: 13.1 ± 1.2 g/dl
Lee 2014	15 randomised	15 randomised
	IRLS score: 21.9 ± 6.01	IRLS score: 21.9 ± 6.25
	1 male (6.7%) Age: 53.3 ± 13.05 years Serum ferritin: 35.5 ± 11.62 µg/l Hb: 13.0 ± 0.80 g/dl	0 male (0.0%) Age: 59.1 ± 10.83 years Serum ferritin: 36.6 ± 7.11 µg/l Hb: 13.0 ± 1.39 g/dl
Trenkwalder 2017	59 randomised	51 randomised
	IRLS score: 25.9 ± 5.65	IRLS score: 26.0 ± 5.78
	11 males (18.6%) Age: 53.0 ± 15.7 years Serum ferritin: 41.93 ± 34.55 µg/l Hb: n.r.	9 males (17.6%) Age: 55.5 ± 15.9 years Serum ferritin: 48.85 ± 45.95 µg/l Hb: n.r.
Wang 2009	11 randomised	7 randomised
	IRLS score: 24.8 ± 5.72	IRLS score: 23.0 ± 5.03
	5 males (45.5%) Age: 60 years (36 – 82)*** Ferritin: 40.6 ± 15.3 ng/ml Hb: 14.5 ± 1.3 g/dl	2 males (28.6%) Age: 58 years (33 – 72) *** Ferritin: 36.7 ± 20.8 ng/ml Hb: 13.7 ± 1.5 g/dl

*data are shown as mean ± standard deviation, unless otherwise specified; **n Cho 2016, baseline serum ferritin values were from screening tests and not from day 1 of the RCT; ***mean (range);****In Earley 2009, units were not reported for the baseline measurements, but it was assumed based on the reporting that they were age in years, serum ferritin in ng/ml and haemoglobin in g/dl. Abbreviations: n.r., not reported; IRLS (by Allen 2011, Grote 2009, Lee 2014), International Restless Legs Study Group Rating

Scale; IRLS (by Wang 2009), International Restless Legs Scale; IRLSS (by Cho 2016 and Earley 2009), International Restless Legs Syndrome Severity Scale; Hb, haemoglobin concentration; RLS, restless legs syndrome;

2.3.2.3 Risk of bias

The risk of selection bias (due to inappropriate random sequence generation) was unclear in five RCTs (Allen 2011, Davis 2000, Earley 2009, Grote 2009, Lee 2014), because the method of the random sequence generation was not reported; and low in three RCTs (Cho 2016, Trenkwalder 2017, Wang 2009), due to adequate random sequence generation. The risk of selection bias (allocation concealment) was unclear in three RCTs (Earley 2009, Lee 2014, Wang 2009), due to insufficient reporting and low in five RCTs (Allen 2011, Cho 2016, Davis 2000, Grote 2009, Trenkwalder 2017), because concealment methods were sufficiently described. The risk of performance bias was high in one RCT (Lee 2014) and low in the seven other RCTs. The risk of detection bias was unclear in five RCTs (Allen 2011, Cho 2016, Davis 2000, Grote 2009, Lee 2014) and low in three RCTs (Earley 2009, Trenkwalder 2017, Wang 2009). The risk of attrition bias for continuous outcome data was high in four RCTs (Allen 2011, Davis 2000, Lee 2014, Trenkwalder 2017), unclear in two RCTs (Cho 2016, Earley 2009) and low in two RCTs (Grote 2009, Wang 2009), while the risk of attrition bias for binary outcome data was high in five RCTs (Allen 2011, Davis 2000, Grote 2009, Lee 2014, Trenkwalder 2017), unclear in two RCTs (Cho 2016, Earley 2009) and low in one RCT(Wang 2009). Reporting bias was high in two RCTs (Earley 2009, Trenkwalder 2017), unclear in five RCTs (Allen 2011, Cho 2016, Davis 2000, Lee 2014, Wang 2009) and low in one RCT (Grote 2009). A summary of the risk of bias assessment is shown in Table 6 and a detailed description with support of judgment can be found in Appendix 3.

Table 6 Adults with restless legs syndrome: Risk of bias

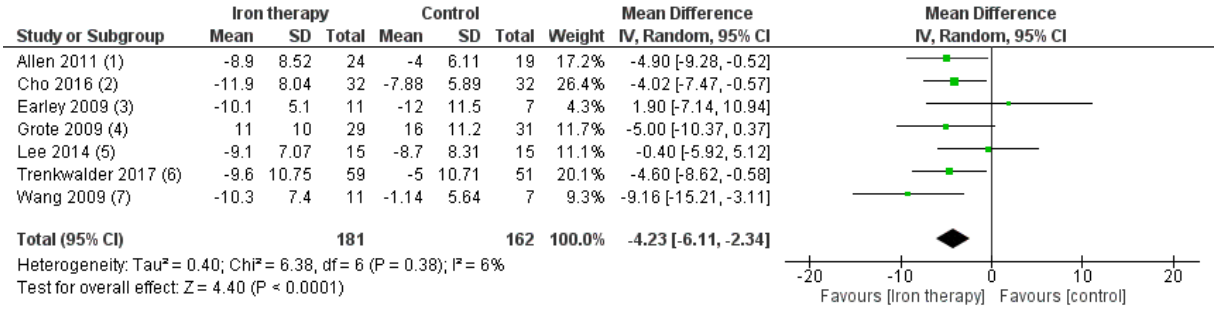
Trial name	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete continuous outcome data (attrition bias)	Incomplete binary data (attrition bias)	Selective reporting (reporting bias)
Allen 2011	Unclear	Low	Low	Unclear	High	High	Unclear
Cho 2016	Low	Low	Low	Unclear	Unclear	Unclear	Unclear
Davis 2000	Unclear	Low	Low	Unclear	High	High	Unclear
Earley 2009	Unclear	Unclear	Low	Low	Unclear	Unclear	High
Grote 2009	Unclear	Low	Low	Unclear	Low	High	Low
Lee 2014	Unclear	Unclear	High	Unclear	High	High	Unclear
Trenkwalder 2017	Low	Low	Low	Low	High	High	High
Wang 2009	Low	Unclear	Low	Low	Low	Low	Unclear

2.3.2.4 Critical outcomes

2.3.2.4.1 Restless legs syndrome symptom severity

Seven RCTs (Allen 2011, Cho 2016, Earley 2009, Grote 2009, Lee 2014, Trenkwalder 2017, Wang 2009) reported on RLS symptom severity with a range of follow-up from two to 12 weeks. All RCTs used the International Restless Legs Syndrome Study Group severity scale (IRLS, range 0 [less severe RLS symptoms] to 40 [more severe RLS symptoms]). However, the RCT authors called the IRLS-Instrument slightly different (IRLS Study Group severity scale, IRLS severity scale, IRLS Group Rating Scale, IRLS Study Group rating scale for severity, IRLS symptoms severity score or IRLS survey, see also Figure 1), but authors referred to the same references from the International Restless Legs Syndrome Group[48-50]. Additional information on the IRLS provided by authors was consistent across those reporting details on the IRLS; therefore, the tools were pooled across all RCTs without standardisation.

Compared to control, iron therapy statistically significantly reduced symptom severity from RLS (MD -4.23, 95% CI [-6.11, -2.34], Figure 2; low quality of evidence, Table 22). Heterogeneity between RCTs was low (I^2=6%). Removing the trial by Lee 2014, which did not use a placebo comparator (comparator: pramipexole) from the analysis did not substantially change the effect estimate (see section 2.3.2.7 Sensitivity analysis).



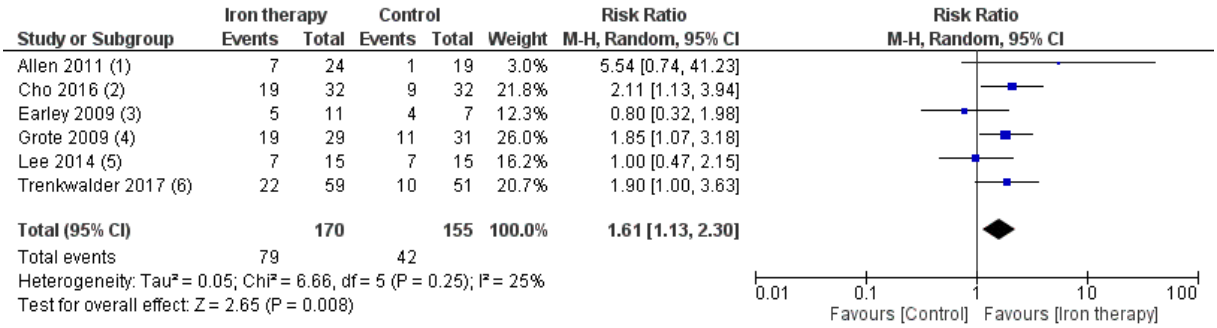
Footnotes
(1) IRLS study group severity scale; range 0-40; change from baseline, negative value indicates improvement; 4 weeks
(2) IRLS Severity scale; range 0-40; change from baseline, negative value indicates improvement; 6 weeks
(3) IRLS study group severity scale; range 0-40; change from baseline, negative value indicates improvement; 2 weeks
(4) IRLS Group Rating Scale; range 0-40, higher score representing worse symtomatic; 11 weeks
(5) IRLS study group rating scale for severity; range 0-40; change from baseline, negative value indicates improvement; 12 weeks
(6) IRLS symptoms severity score; range 0-40; SD calculated from SE; change from baseline, negative value indicates improvement; 12 weeks
(7) IRLS survey; range 0-40; change from baseline, negative value indicates improvement; 12 weeks

Figure 2 Adults with RLS, symptom severity of RLS

2.3.2.4.2 Restless legs syndrome treatment response

Six RCTs (Allen 2011, Cho 2016, Earley 2009, Grote 2009, Lee 2014, Trenkwalder 2017) reported on RLS treatment response with a range of follow-up from two to 12 weeks. Three RCTs (Grote 2009, Lee 2014, Trenkwalder 2017) defined treatment responders as ≥50% reduction on the International Restless Legs Syndrome Study Group severity scale (IRLS). One RCT (Cho 2016) defined treatment responders as ≥40% decrease on the IRLS and one RCT (Earley 2009) defined treatment responders as improvements sufficient enough not to go back on any RLS medication. One RCT (Allen 2011) defined RLS remitters as ≤10 IRLS score at four weeks. When pooling all trials with these differently defined endpoints, iron therapy showed a statistically significant increase in RLS treatment response when compared to control (RR 1.61, 95% CI [1.13, 2.30], Figure 3; very low quality of evidence, Table 22). Heterogeneity between the RCTs was low (I^2=25%). Removing the trial of Lee 2014, which did not use

a placebo comparator (comparator: pramipexole) from the analysis did not substantially change the effect estimate (see 2.3.2.7 Sensitivity analysis).

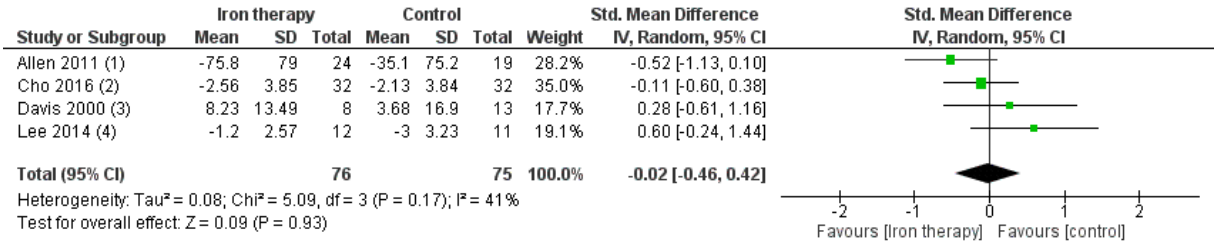


Footnotes
(1) RLS remitters: ≤10 IRLS score; 4 weeks
(2) Treatment responders: > 40% IRLSS scale decrease; 6 weeks
(3) Treatment responders: improvements sufficient enough not to go back on any RLS medication; 2 weeks
(4) Treatment responders: ≥50% IRLS reduction; 11 weeks
(5) Treatment responders: ≥50% IRLS reduction; 12 weeks
(6) Treatment responders: ≥50% IRLS reduction; Assumed 12 weeks

Figure 3 Adults with RLS, treatment response

2.3.2.4.3 Sleep

Four RCTs (Allen 2011, Cho 2016, Davis 2000, Lee 2014) reported on sleep quality with a range of follow-up from four to 12 weeks. Two RCTs (Cho 2016, Lee 2014) used the Pittsburg Sleep Quality Index (PSQI, range from 0 to 21 [higher value indicating worse sleep quality]). One RCT (Allen 2011) used the Medical Outcome Study sleep scale (MOS, [higher score indicating better sleep]) and one RCT (Davis 2000) used a Visual Analog Scale (VAS, range from 0 [impossible to sleep] to 100 [slept very well]). Compared to control, iron therapy had no statistically significant effect on sleep (SMD -0.02, 95% CI [-0.46, 0.42], Figure 4; very low quality of evidence, Table 22). Heterogeneity between the RCTs was moderate (I^2=41%). In the sensitivity analyses, when the trial by Lee 2014 was excluded, the heterogeneity decreased to 12% (see Figure 17). Lee 2014 was the only RCT using pramipexole (dopamine agonist) as comparator instead of placebo. It is known that dopamine agonists have a slight effect on improving sleep quality in patients with RLS[51].



Footnotes
(1) MOS sleep total score; range n.r.; change from baseline, neg. score indicates better sleep; multiplied with -1 to invert effects; 4 weeks
(2) Pittsburg Sleep Quality Index (PSQI); range 0-21; change from baseline, negative value indicates improved sleep quality; 6 weeks
(3) VAS, Quality of Sleep; range 0-100; change from baseline, positive value indicates worse sleep quality; multiplied with -1 to invert effect; 12 weeks
(4) Pittsburg Sleep Quality Index (PSQI); range 0-21; change from baseline, negative value indicates improved sleep quality; 12 weeks

Figure 4 Adults with RLS, sleep

2.3.2.4.4 Sleepiness

One RCT (Lee 2014) used the Epworth Sleepiness Scale (ESS, range 0 to 24 [higher score indicating higher daytime sleepiness]). Iron therapy did not show a statistically significant difference when compared to control on sleepiness (MD -0.20, 95% CI [-2.66, 2.26], Figure 5; very low quality of evidence, Table 22).

An additional RCT (Trenkwalder 2017) reported Daytime Tiredness (1 item of the 6 items of the Restless Legs Syndrome-6 rating scale), range from 0 no symptoms to 10 very severe symptoms, and found a statistically significant effect (least-squares MD -1.5, 95% CI [-2.47, -0.56]) for iron therapy when compared to placebo.

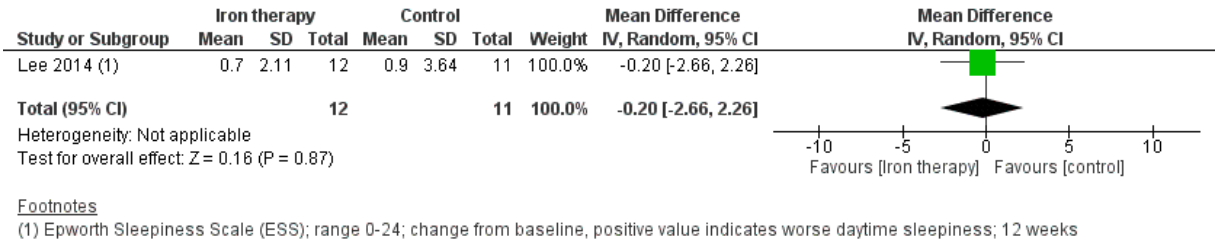


Figure 5 Adults with RLS, sleepiness

2.3.2.4.5 Adverse events

Three RCTs (Allen 2011, Grote 2009, Trenkwalder 2017) reported the number of patients reporting adverse events. When comparing iron therapy to control, the risk for adverse events was not statistically significantly increased (RR 1.37, 95% CI [0.88, 2.13], Figure 29). Heterogeneity between the RCTs was low (I^2=0%).

Davis 2000 and Earley 2009 only reported the number of adverse events and side effects, respectively. Davis 2000 and Earley 2009 were therefore not pooled with the three RCTs (Allen 2011, Grote 2009, Trenkwalder 2017). Davis 2000 reported 12 adverse events at 14 weeks follow-up in the iron therapy group in a total of 14 randomised patients and zero adverse events in the placebo group in a total of 14 randomised patients. Earley 2009 reported at the day of infusion 13 side effects in the iron therapy group in a total of 11 randomised patients and two side effects in the placebo group in a total of 7 randomised patients. All side effects resolved within minutes or hours after infusion. Earley 2009 described no adverse outcomes in the iron therapy group and placebo group at two weeks follow-up.

Additional results on adverse events can be found in section 2.3.5.1.1

Importantly, the type of adverse event was insufficiently reported by the individual RCTs and was therefore not listed within this section. More information can be found in section 2.5.4.

2.3.2.4.6 Serious adverse events

Serious adverse events were reported in four RCTs (Allen 2011, Cho 2016, Davis 2000, Trenkwalder 2017). Across the four RCTs, only a total of two serious adverse events were reported. Davis 2000 reported one vertebral fracture, while Trenkwalder 2017 did not specify the reported serious adverse event. The remaining two trials reported no serious adverse events until the latest follow-up. When comparing iron therapy with control, the risk for serious adverse events was not statistically significantly increased (RR 2.85, 95% CI [0.31, 26.38], Figure 30). Heterogeneity between the RCTs was

low (I^2=0%). Mortality was explicitly mentioned in three RCTs with no deaths reported (Earley 2009, Trenkwalder 2017, Wang 2009).

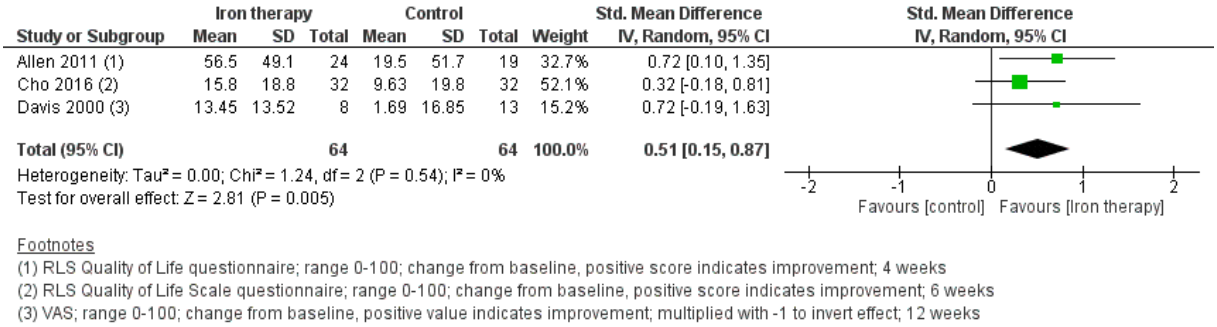
Additional results on serious adverse events can be found in section 2.3.5.1.2.

2.3.2.5 Important outcomes

2.3.2.5.1 Quality of life

Quality of life was reported in four RCTs (Allen 2011, Cho 2016, Davis 2000, Wang 2009) with a range of follow-up from four to 12 weeks. Three RCTs (Allen 2011, Cho 2016, Davis 2000) reported continuous outcomes, which were pooled. Two RCTs (Allen 2011, Cho 2016) used the Restless Legs Syndrome Quality of Life questionnaire (RLS QoL, range 0 to 100 [higher score indicating a better quality of life]. The third RCT used a Visual Analog Scale (VAS, range 0 [does not affect my life] to 100 [makes my life miserable]). There was a statistically significant effect in favour of iron therapy compared to control (SMD 0.51, 95% CI [0.15, 0.87], Figure 6; very low quality of evidence, Table 22). Heterogeneity between RCTs was low (I^2=0%).

Wang 2009 reported at 12 weeks follow-up whether quality of life had "improved" or "stayed the same or worsened" (binary outcome). The improved quality of life in seven of 11 participants in the iron therapy group compared to one of seven participants in the placebo group was reported to be not statistically significant (P=0.07).

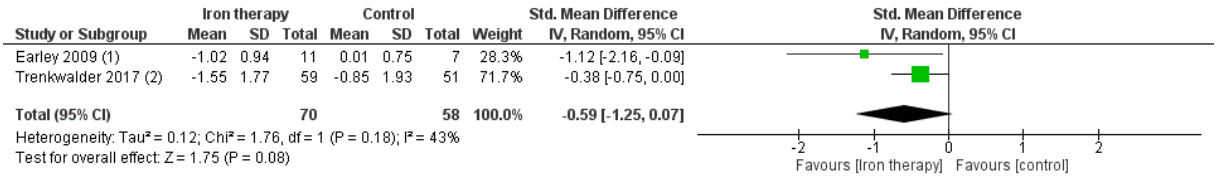


Footnotes
(1) RLS Quality of Life questionnaire; range 0-100; change from baseline, positive score indicates improvement; 4 weeks
(2) RLS Quality of Life Scale questionnaire; range 0-100; change from baseline, positive score indicates improvement; 6 weeks
(3) VAS; range 0-100; change from baseline, positive value indicates improvement; multiplied with -1 to invert effect; 12 weeks

Figure 6 Adults with RLS, quality of life

2.3.2.5.2 Global impression rating

Global impression rating was reported for two RCTs (Earley 2009, Trenkwalder 2017) with a range of follow-up from two to 12 weeks. One RCT (Earley 2009) used the Global Rating Scale (GRS, range 0 [no symptoms] to 6 [very severe symptoms]) and one RCT (Trenkwalder 2017) used the Clinical Global Impression – Item 1 (CGI-1, seven point severity scale [higher value indicating worse severity]). There was no statistically significant effect on global impression rating in favour of iron therapy compared to control (SMD -0.59, 95% CI [-1.25, 0.07], Figure 6; very low quality of evidence, Table 22). Heterogeneity between the RCTs was moderate (I^2=43%).



Footnotes
(1) GRS; range from no symptoms [0] to very severe symptoms [6]; change from baseline, neg. value indicate improvement; 2 weeks
(2) Clinical Global Impression - Item 1; range n.r. , 7 point severity scale; least squares mean change from baseline; SD calculated from SE, estimated from graph; 12 weeks

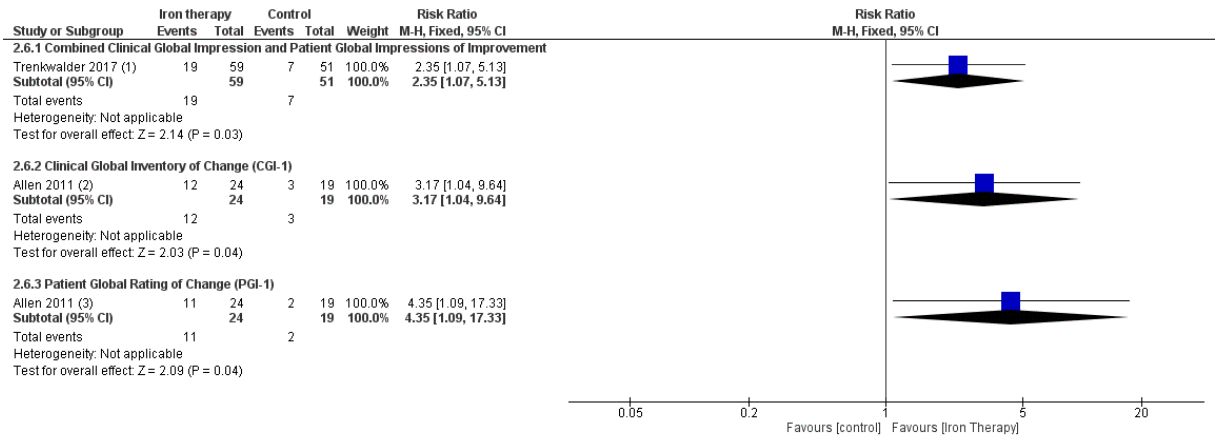
Figure 7 Adults with RLS, global impression rating

2.3.2.5.3 Change in global impression, inventory or rating

Global impression improvement was reported in two RCTs (Allen 2011, Trenkwalder 2017) with a range of follow-up from four to 12 weeks. Trenkwalder 2017 combined the Clinical Global Impression – Item 2 (CGI-Item 2) and Patient Global Impressions of Improvement (PGI-1) to assess the number of patients with ratings of much improvement in CGI-Item 2 and very much improvement in PGI-1 item. When comparing iron therapy with control, a statistically significant effect of iron therapy on change in global impression was found (RR 2.35, 95% CI [1.07, 5.13], Figure 8; very low quality of evidence, Table 22).

Allen 2011 used the Clinical Global Inventory of Change (CGI-1) to assess the number of patients with very much or much improved change in global impression, and found a statistically significant effect in favour of iron therapy when compared to control (RR 3.17, 95% CI [1.04, 9.64], Figure 8; very low quality of evidence, Table 22).

Allen 2011 also used the Patient Global Rating of Change (PGI-1) to assess the number of patients with very much or much improved change in global impression, and found a statistically significant effect in favour of iron therapy when compared to control (RR 4.35, 95% CI [1.09, 17.33], Figure 8; very low quality of evidence, Table 22).

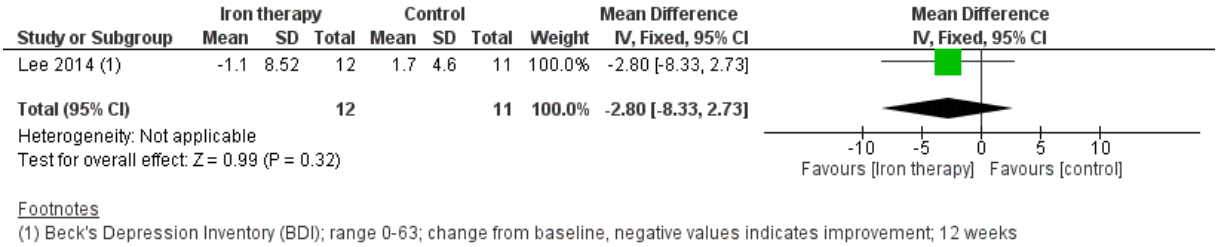


Footnotes
(1) Clinical Global Impression - Item 2 and Patient Global Impressions of Improvement (combined) ; n of patients with a rating of much improved in CGI - Item 2 and very much improved in PGI - 1; 12 weeks
(2) Clinical Global Inventory of Change (CGI-1); n of patients that were very much or much improved; n patients extrapolated from percentages; 4 weeks
(3) Patient Global Rating of Change (PGI-1); n of patients that were very much or much improved; n patients extrapolated from percentages; 4 weeks

Figure 8 Adults with RLS, change in global impression, inventory or rating improvement

2.3.2.5.4 Depression

One RCT (Lee 2014) reported on depression at 12 weeks follow-up using the Beck's Depression Inventory (BDI, range 0 to 63 [higher score indicating worse depression]). Iron therapy compared to control had no statistically significant effect on depression (MD -2.80, 95% CI [-8.33, 2.73], Figure 9; very low quality of evidence, Table 22).

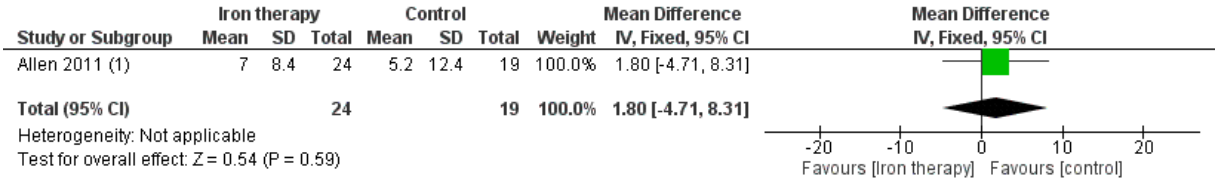


Footnotes
(1) Beck's Depression Inventory (BDI); range 0-63; change from baseline, negative values indicates improvement; 12 weeks

Figure 9 Adults with RLS, depression

2.3.2.5.5 Fatigue

One RCT (Allen 2011) reported fatigue at four weeks follow-up using the Fatigue Severity Scale (FSS, range 9 to 63 [higher score indicating higher level of fatigue]). There was no statistically significant difference between iron therapy and control (MD 1.80, 95% CI [-4.71, 8.31], Figure 10; very low quality of evidence, Table 22)



Footnotes
(1) Fatigue Severity Scale (FSS), probably reported as total score; range 9-63; change from baseline, neg. value indicates improvement; 4 weeks

Figure 10 Adults with RLS, fatigue

2.3.2.6 Subgroup analyses

Seven RCTs reported on RLS severity and six on RLS treatment responses. For both outcomes, there was no heterogeneity measured. The subgroups female vs. male and adolescents/children vs. adults were not analysed because information for these subgroups was lacking. Two of the pre-specified subgroup analyses have been considered "of interest", and are presented here. However, because of the limited number of RCTs per subgroup, differences between subgroups need to be interpreted with care.

2.3.2.6.1 Subgroup 1: Oral vs. intravenous therapy with iron

2.3.2.6.1.1 Restless legs syndrome symptom severity, Subgroup 1: Oral vs. intravenous therapy

Of the seven RCTs reporting on RLS severity, five RCTs administered iron intravenously and two orally. Of the RCTs with oral iron application, one RCT (Lee 2014) used pramipexole as comparator and was, therefore, analysed as a separate subgroup. The test for subgroup differences was not statistically

significant (P = 0.11). Too few RCTs administrating oral iron reported this outcome and no conclusions for oral vs. intravenous iron application on RLS symptom severity can be made.

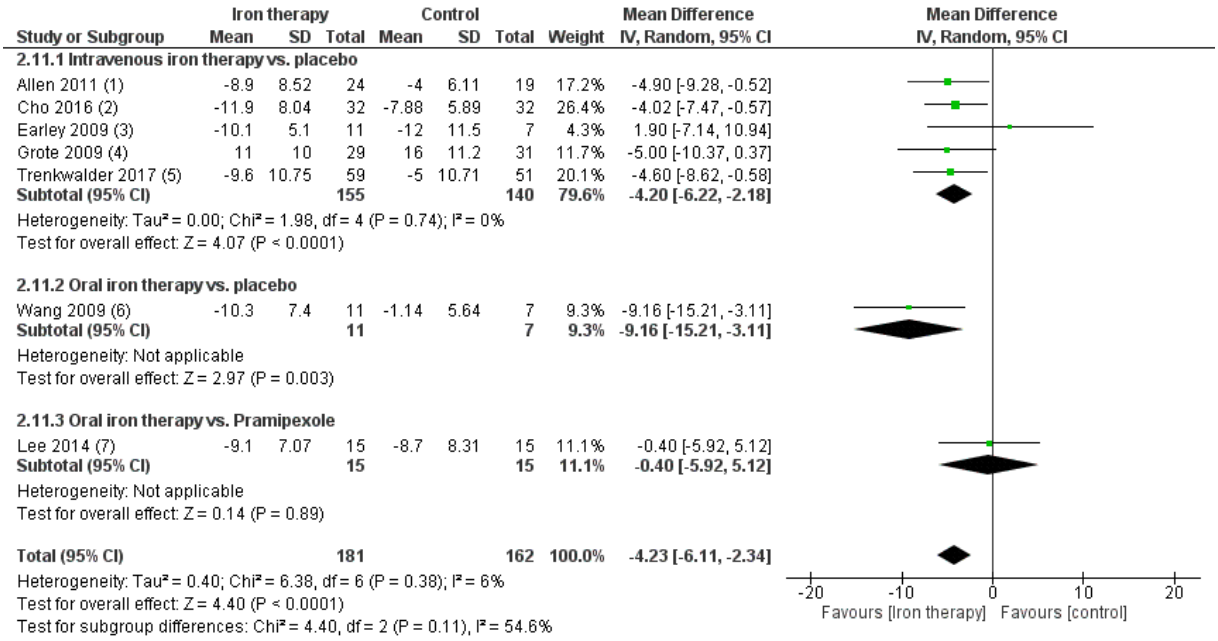


Figure 11 RLS symptom severity, subgroup: route of administration

2.3.2.6.1.2 Restless legs syndrome treatment response, Subgroup 1: Oral vs. intravenous therapy

Of the six RCTs reporting on RLS response, five placebo-controlled RCTs had administrated iron intravenously and one orally (Wang 2009). One RCT (Lee 2014) administered iron orally and with pramipexole being the comparator. The test for subgroup differences was not statistically significant (P = 0.18). Too few RCTs with oral iron administration were available and hence no conclusions regarding oral vs. intravenous iron therapy on RLS treatment response can be made.

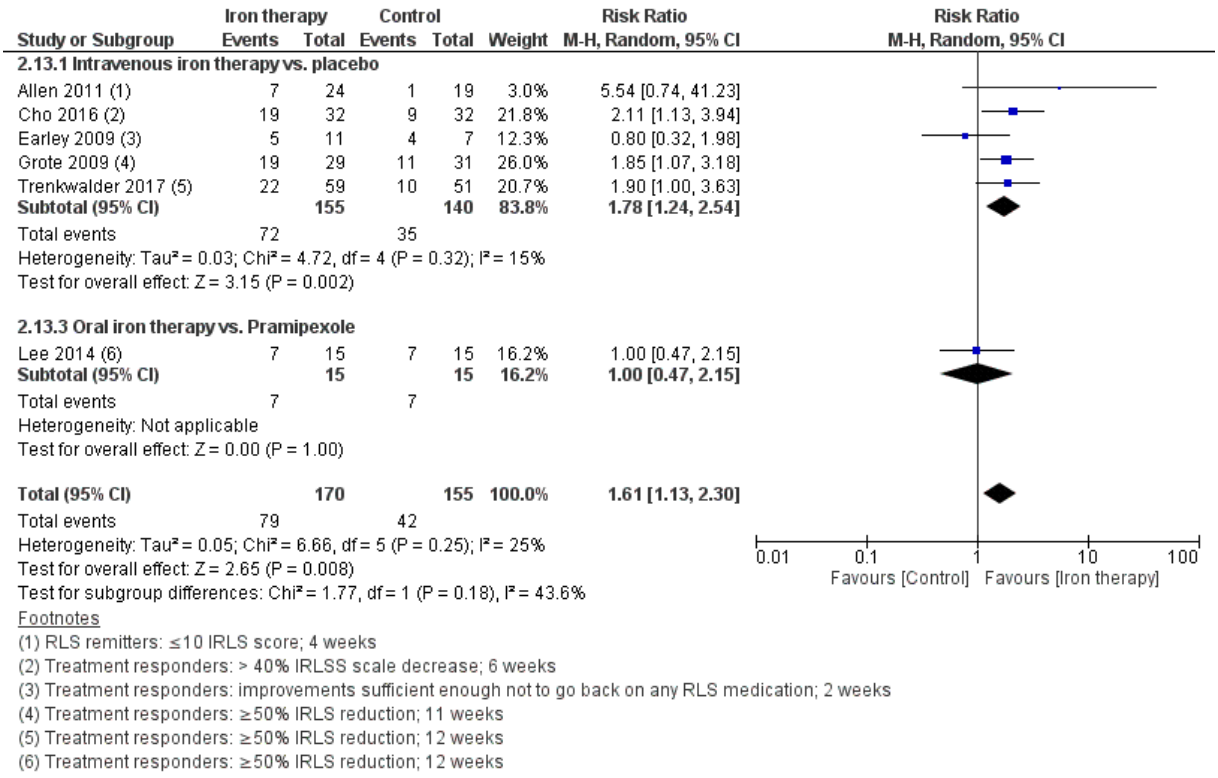


Figure 12 RLS treatment response, subgroup: route of administration

2.3.2.6.2 Subgroup 2: Iron status of study population at recruitment

2.3.2.6.2.1 Restless legs syndrome symptom severity, Subgroup 2: Iron status at recruitment

Of the seven RCTs reporting on RLS severity, in three RCTs the iron status of the study population was unclear, in two RCTs the patient population had a mixed iron status (low and normal serum ferritin concentrations, i.e. ≤75 µg/l), in two RCTs the study populations had serum ferritin concentrations below 50 µg/l, with pramipexole being the comparator in one RCT. The test for subgroup differences was not statistically significant (P = 0.41).

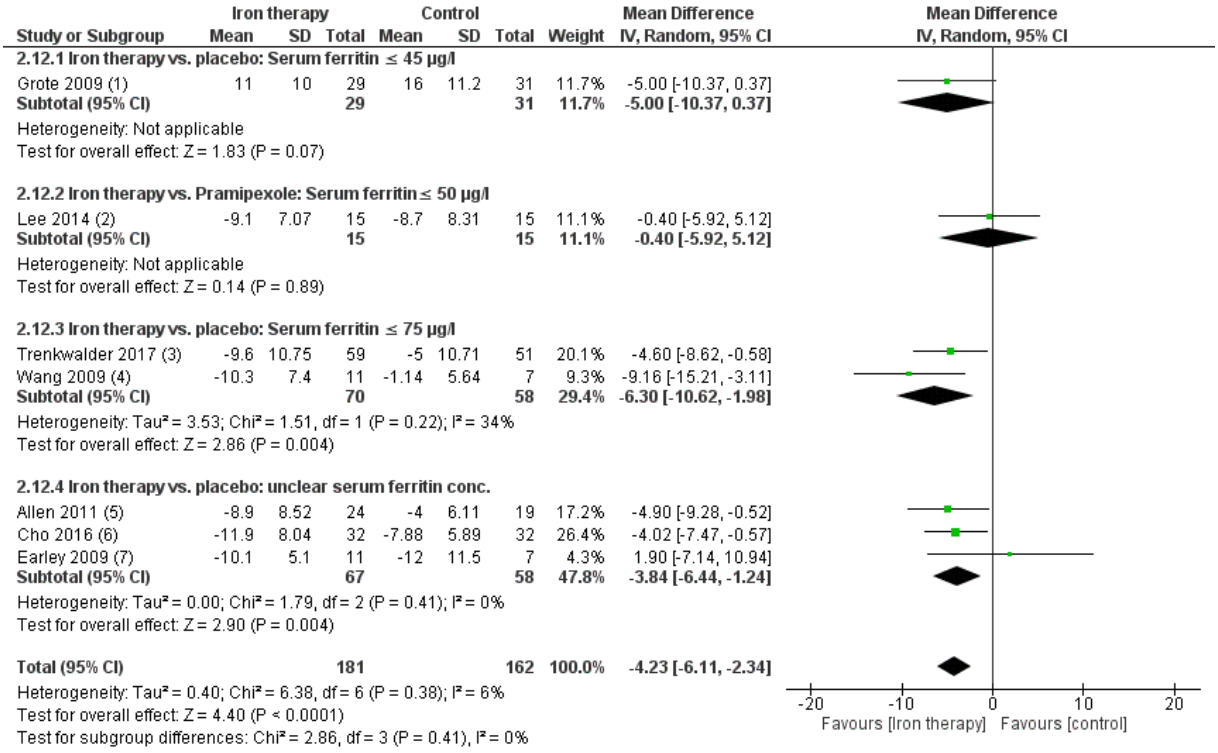


Figure 13 RLS symptom severity, subgroup: iron status of the study population at recruitment

2.3.2.6.2.2 Restless legs syndrome treatment response, Subgroup 2: Iron status at recruitment

Of the six RCTs reporting on RLS treatment response, in three RCTs the study populations had an unclear iron status, one RCT population had a mixed iron status (low and normal serum ferritin concentrations, i.e. ≤75 µg/l), in two RCTs the study populations had serum ferritin concentrations below 50 µg/l, with pramipexole being the comparator in one RCT. The test for subgroup differences was not statistically significant (P = 0.57).

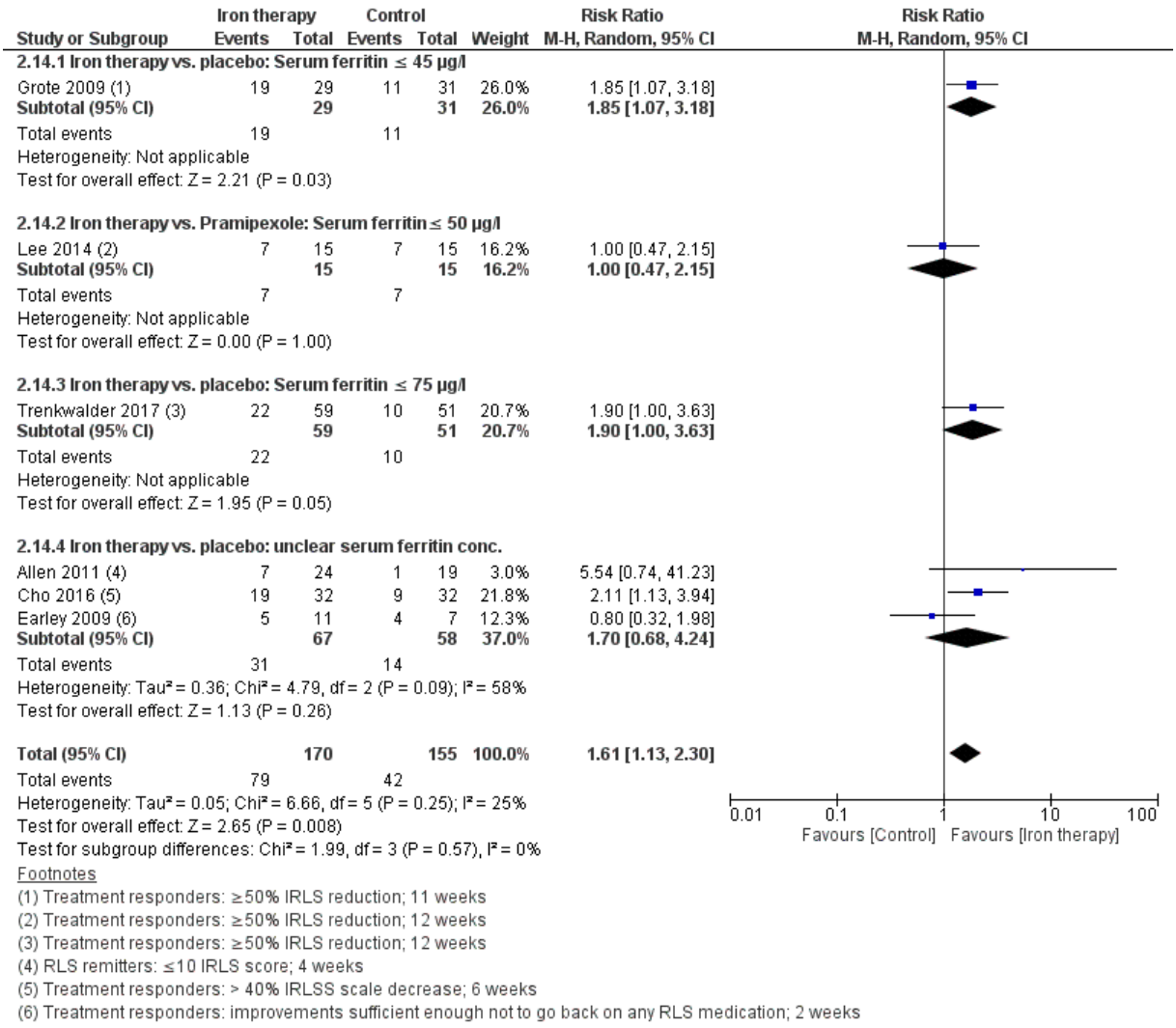
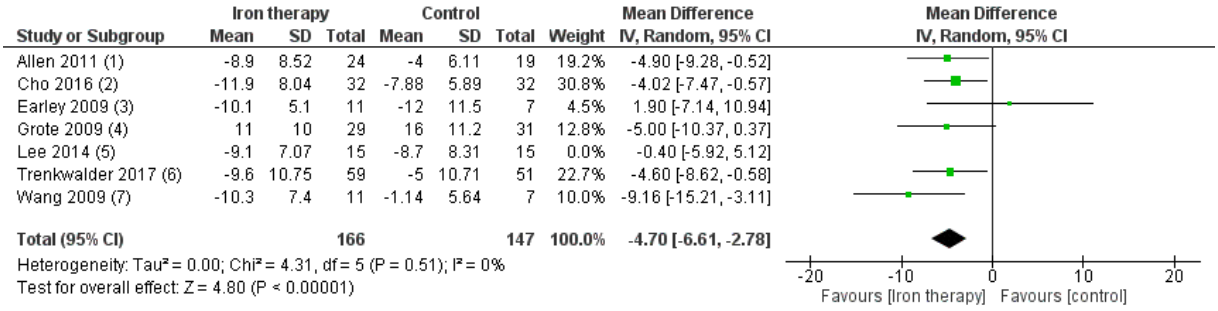


Figure 14 RLS treatment response, subgroup: iron status of the study population at recruitment

2.3.2.7 Sensitivity analyses

Sensitivity analyses were conducted excluding the trial by Lee 2014, which used pramipexole as comparator.

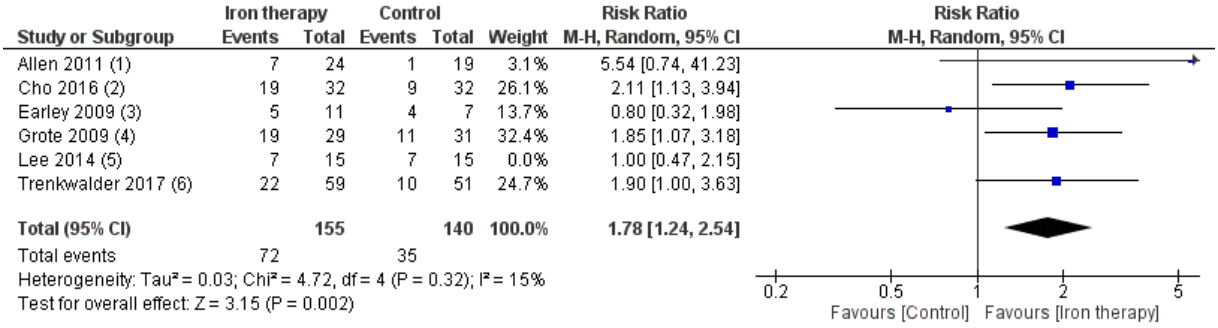
Excluding Lee 2014 only marginally changed the effects of iron therapy on the outcomes of RLS symptom severity and RLS treatment response (Figure 15, Figure 16). Excluding the trial by Lee 2014 for the outcome of sleep moderately reduced the high heterogeneity from I^2=41% to 12% with a slight change in the effect estimate (Figure 17).



Footnotes

(1) IRLS study group severity scale; range 0-40; change from baseline, negative value indicates improvement; 4 weeks
(2) IRLS Severity scale; range 0-40; change from baseline, negative value indicates improvement; 6 weeks
(3) IRLS study group severity scale; range 0-40; change from baseline, negative value indicates improvement; 2 weeks
(4) IRLS Group Rating Scale; range 0-40, higher score representing worse symtomatic; 11 weeks
(5) IRLS study group rating scale for severity; range 0-40; change from baseline, negative value indicates improvement; 12 weeks
(6) IRLS symptoms severity score; range 0-40; SD calculated from SE; change from baseline, negative value indicates improvement; 12 weeks
(7) IRLS survey; range 0-40; change from baseline, negative value indicates improvement; 12 weeks

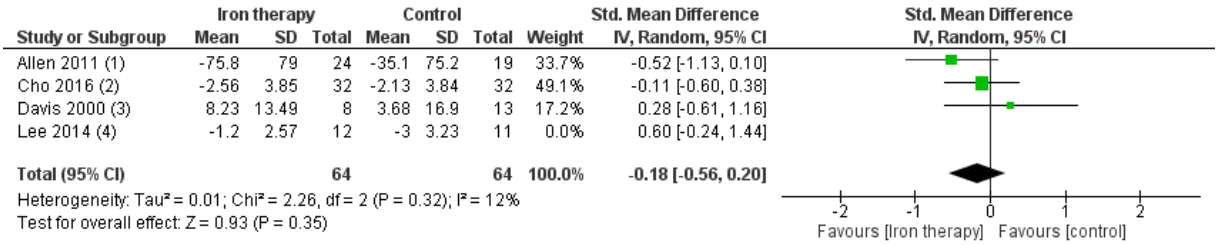
Figure 15 Sensitivity analyses: RLS symptom severity without Lee 2014



Footnotes

(1) RLS remitters: ≤10 IRLS score; 4 weeks
(2) Treatment responders: > 40% IRLSS scale decrease; 6 weeks
(3) Treatment responders: improvements sufficient enough not to go back on any RLS medication; 2 weeks
(4) Treatment responders: ≥50% IRLS reduction; 11 weeks
(5) Treatment responders: ≥50% IRLS reduction; 12 weeks
(6) Treatment responders: ≥50% IRLS reduction; 12 weeks

Figure 16 Sensitivity analyses: RLS treatment response without Lee 2014



Footnotes

(1) MOS sleep total score; range n.r.; change from baseline, neg. score indicates better sleep; multiplied with -1 to invert effects; 4 weeks
(2) Pittsburg Sleep Quality Index (PSQI); range 0-21; change from baseline, negative value indicates improved sleep quality; 6 weeks
(3) VAS, Quality of Sleep; range 0-100; change from baseline, positive value indicates worse sleep quality; multiplied with -1 to invert effect; 12 weeks
(4) Pittsburg Sleep Quality Index (PSQI); range 0-21; change from baseline, negative value indicates improved sleep quality; 12 weeks

Figure 17 Sensitivity analyses: sleep without Lee 2014

2.3.3 Women with fatigue

The RCT characteristics and risk of bias assessment in women with fatigue are presented first followed by the presentation of the results for each outcome.

2.3.3.1 Overview of included RCTs

Six references (four full research articles and two abstracts) encompassing four relevant RCTs were identified. References can be found in Table 7. An overview of the included outcomes with analysed follow-up time-points from each RCT is given in Table 8.

Table 7 Women with fatigue: Overview of included RCTs, their study names and references

Trial name	Reference (Main reference highlighted in colour)
FERRIM (Krayenbuehl 2011)[52]	Krayenbuehl PA, Battegay E, Breymann C, Furrer J, Schulthess G. Intravenous iron for the treatment of fatigue in nonanemic, premenopausal women with low serum ferritin concentration. *Blood.* 2011;118(12):3222-3227.
PREFER (Favrat 2014)[53-55]	Favrat B, Balck K, Breymann C, et al. Evaluation of a single dose of ferric carboxymaltose in fatigued, iron-deficient women--PREFER a randomized, placebo-controlled study. *PloS one.* 2014;9(4):e94217.
	Favrat B, Balck K, Gasche C, et al. A single 1000mg iron dose of ferric carboxymaltose improves fatigue in iron deficient, non-anaemic premenopausal women - Results of the randomised, placebo-controlled prefer study. *International journal of gynaecology and obstetrics.* 2012;119:S858-s859.
	Favrat B, Balck K, Gasche C, et al. One 1000 mg iron dose of ferric carboxymaltose improved fatigue in iron-deficient, non-anaemic women in the randomised placebo-controlled study PREFER. *Bjog.* 2012;119:232-233.
Vaucher 2012[56]	Vaucher P, Druais PL, Waldvogel S, Favrat B. Effect of iron supplementation on fatigue in nonanemic menstruating women with low ferritin: a randomized controlled trial. *Cmaj.* 2012;184(11):1247-1254.
Verdon 2003[9]	Verdon F, Burnand B, Stubi CL, et al. Iron supplementation for unexplained fatigue in non-anaemic women: double blind randomised placebo controlled trial. *Bmj.* 2003;326(7399):1124.

Table 8 Women with fatigue: Overview of the included outcomes with analysed follow-up time-points

Outcome / Trial name	Fatigue Severity	Fatigue Improvement	Quality of Life	Depression	Anxiety	Adverse Events	Serious Adverse Events
FERRIM (Krayenbuehl 2011)	12	12				12	12
PREFER (Favrat 2014)	8	8	8			8	8
Vaucher 2012	12		12	12	12		
Verdon 2003	4			4	4		

The numbers in the fields denote the analysed follow-up period in weeks.

2.3.3.2 Characteristics of the included RCTs

General characteristics of RCTs in women with fatigue are summarised in Table 9 and selected baseline characteristics of the patients from each RCT are presented in Table 10. Four RCTs in women with fatigue were identified and all RCTs were multicentre RCTs conducted in Europe. The four RCTs included 726 non-anaemic women with fatigue and iron deficiency. Length of follow-up ranged from 4 weeks to 12 weeks and all RCTs used placebo as the comparator. In the two RCTs, FERRIM and PREFER, women received parenteral iron therapy and in the other two RCTs, Verdon 2003 and Vaucher 2012, women received oral iron therapy. All four RCTs were industry-sponsored.

Table 9 Women with fatigue: Summary of RCT characteristics and intervention

Trial name **Country**	**Setting** **Enrollment period** **Time-points of FU**	**Population** **Key inclusion criteria***	**Intervention** Compound Dosage regimen	**Comparator** Compound Dosage regimen
FERRIM (Krayenbuehl 2011) Switzerland	Multicentre (4 sites) n.r. 6 and 12 weeks	Women with symptomatic fatigue, iron deficiency and no anaemia Fatigue: "women presenting with fatigue" ID: SF ≤50ng/ml and TSAT ≤50% Anaemia: Hb: ≥12.0 g/dl	Ferric sucrose, prolonged-release (Venofer®; Vifor Pharma) Intravenous, 4 doses containing 200 mg in 200 ml saline solution at 4 days during first two weeks	Placebo Intravenous, 4x200 ml at 4 days during first two weeks
PREFER (Favrat 2014) AT, CH, DE, SE	Multicentre (21 sites) Nov. 2010 – Nov. 2011** 1, 4 and 8 weeks	Women with symtomatic fatigue, iron deficiency and no anaemia Fatigue: ≥5 points on PFS ID: SF <15 µg/l OR [SF <50 µg/l and TSAT <20%] Hb ≥11.5 g/dl	Ferric caboxymaltose Intravenous, single-dose, 1000 mg in 250 ml saline solution at Day 0	Placebo Intravenous, Single dose 250 ml NaCl solution at Day 0
Vaucher 2012 France	Multicentre (44 sites) Apr. 2006 – Aug. 2006 6 and 12 weeks	Women with considerable fatigue, iron deficiency and no anaemia Fatigue: >6 on a 1 – 10 Likert scale ID: SF <50 µg/l Anaemia: Hb ≥12.0 g/dl	Ferrous sulfate (Tardyferon®; Pierre Fabre Médicament) Oral , 80 mg/day for 12 weeks	Placebo Oral, daily for 12 weeks
Verdon 2003 Switzerland	Multicentre (9 sites) Dec. 1997 – Mar. 2000 4 weeks	Women with symptomatic fatigue and no anaemia Fatigue: "women consulting for fatigue" ID: not an inclusion criteria, but 85% of the participants were below SF <50 µg/l Anaemia: Hb ≥11.7 g/l	Ferrous sulphate, long acting (Tardyferon®, Robapharm) Oral, 80 mg/day for 4 weeks	Placebo Oral, daily for 4 weeks

*see Appendix 5.2 for more details on inclusion and exclusion criteria ** Study period

Abbreviations: AT, Austria; CH, Switzerland; DE, Germany; FU, Follow-up; Hb, haemoglobin concentration; ID, iron deficiency; n.r., not reported; PFS, Piper Fatigue Scale; SE, Sweden; SF, serum ferritin; TSAT, transferrin saturation

Table 10 Women with fatigue: Baseline characteristics

Trial name	Intervention Group*	Comparator Group*
FERRIM (Krayenbuehl 2011)	43 randomised Fatigue severity (BFI[a]): 4.0 (n.r)** Age: 31 ± 8 years Serum ferritin: 24 ng/ml (10, 32) ** Hb: 13.3 ± 0.6 g/dl	47 randomised Fatigue severity (BFI[a]): 4.7 (n.r)** Age: 32 ± 7 years Serum ferritin: 20 ng/ml (14, 28) ** Hb: 13.3 ± 0.7 g/dl
PREFER (Favrat 2014)	145 randomised Fatigue severity (PFS[b]): 6.4 (5.7, 7.2)** Age: 34.6 ± 8.8 years Serum ferritin: 15 µg/l (10, 25) ** Hb: 12.8 g/dl (12.4, 13.5)**	149 randomised Fatigue severity (PFS[b]): 6.4 (5.5, 7.3)** Age: 35.0 ± 9.6 years Serum ferritin: 16 µg/l (11, 28) ** Hb: 12.9 g/dl (12.2, 13.4)**
Vaucher 2012	102 randomised Fatigue severity (MAF[c]): 37.4 ± 6.2 Age: 36.4 ± 9.3 years Serum ferritin: 22.5 ± 12.7 µg/l Hb: 13.5 ± 0.9 g/dl	96 randomised Fatigue severity (MAF[c]): 37.0 ± 5.9 37.3 ± 9.5 years Serum ferritin: 23.3 ± 11.6 µg/l Hb: 13.6 ± 0.8 g/dl
Verdon 2003	75 randomised (baseline characteristics reported for n=71) Fatigue severity (VAS[d]): 6.4 ± 1.6 Age: 36.1 ± 9.9 years Serum ferritin: 30.4 ± 31 µg/l Hb: 13.54 ± 0.95 g/dl	69 randomised (baseline characteristics reported for n=65) Fatigue severity (VAS[d]): 6.5 ± 1.6 Age: 34.6 ± 11.5 years Serum ferritin: 29.2 ± 28 µg/l Hb: 13.65 ± 1.04 g/dl

*data are shown as mean ± standard deviation, unless otherwise specified; **median (Q1, Q3)

[a] BFI, Brief Fatigue Inventory, range 0-10, 0 indicates no and 10 maximum imaginable fatigue [b] PFS, Piper Fatigue Scale, range 1-10, 1 indicates no fatigue at all and 10 very severe fatigue [c] MAF, Multidimensional Assessment of Fatigue score, range 0-50, higher score indicate worsening of fatigue [d] VAS, Visual Analog Scale, range 1-10, 1 indicates no fatigue at all and 10 very severe fatigue

Abbreviations: Hb, haemoglobin concentration; n.r., not reported

2.3.3.3 Risk of bias

The risk of selection bias (due to random sequence generation) was considered low in two RCTs (PREFER, Vaucher 2012) because random sequence generation was adequate; and unclear in two RCTs (FERRIM, Verdon 2003), because the method of the random sequence generation was not or insufficiently reported. Allocation concealment was adequate in all RCTs. The risk of performance bias was unclear in one RCT (PREFER); and the risk of performance bias was considered low in the three remaining RCTs since participants and trial staff were adequately blinded. Detection bias was unclear in two RCTs (FERRIM, PREFER) because adequate blinding of outcome assessors was unclear. The risk of attrition bias for continuous outcome data was unclear in one RCT (PREFER); and was low in three RCTs (FERRIM, Vaucher 2012, Verdon 2003), because drop-out rates were low and/or comparable across study groups. The risk of attrition bias for binary outcome data was unclear in two RCTs (PREFER, Vaucher 2012), and judged low for the remaining two RCTs (FERRIM, Verdon 2003). Risk of reporting bias was judged unclear in two RCTs (Vaucher 2012, Verdon 2003); and low in the two remaining (FERRIM, PREFER). A summarised version of the risk of bias assessment is shown in Table 11 and a detailed summary with support of judgment can be found in Appendix 3.

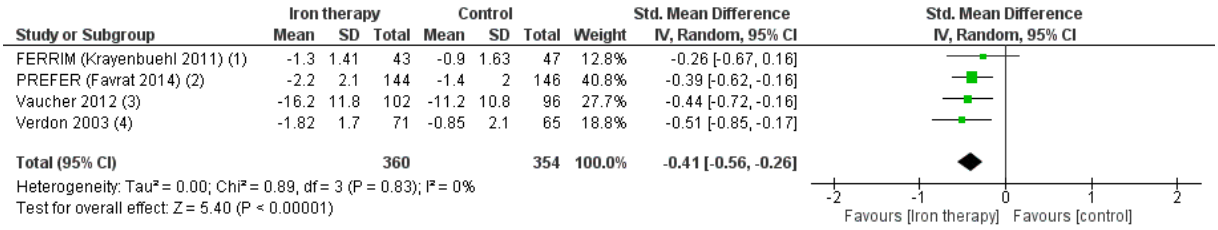
Table 11 Women with fatigue: Risk of bias

Trial name	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment, (detection bias)	Incomplete continuous outcome data (attrition bias)	Incomplete binary data (attrition bias)	Selective reporting (reporting bias)
FERRIM (Krayenbuehl 2011)	Unclear	Low	Low	Unclear	Low	Low	Low
PREFER (Favrat 2014)	Low	Low	Unclear	Unclear	Unclear	Unclear	Low
Vaucher 2012	Low	Low	Low	Low	Low	Unclear	Unclear
Verdon 2003	Unclear	Low	Low	Low	Low	Low	Unclear

2.3.3.4 Critical outcomes

2.3.3.4.1 Fatigue severity

Four RCTs (FERRIM, PREFER, Vaucher 2012, Verdon 2003) reported on fatigue severity with a range of follow-up from four to 12 weeks. Each RCT used a different scale. The FERRIM RCT used the Brief Fatigue Inventory (BFI, range 0 [no fatigue] to 10 [maximum imaginable fatigue]), while the PREFER RCT used the 22-item Piper Fatigue Scale (PFS, range 1 [no fatigue at all] to 10 [very severe fatigue]). Vaucher 2012 used the global fatigue index of the Multidimensional Assessment of Fatigue Scale (MAF, range 0 [less fatigued] to 50 [more fatigued]); and Verdon 2003 used a Visual Analog Scale (VAS, range 1 [no fatigue] to 10 [very severe fatigue]). Compared to placebo, iron therapy showed a statistically significant reduction of fatigue symptom severity (SMD -0.41, 95% CI [-0.56, -0.26], Figure 18; moderate quality of evidence, Table 23). Heterogeneity between RCTs was low (I^2=0%). Further analysis based on individual patient data to assess the association between fatigue severity and ferritin concentration at baseline have been summarized in section 2.3.3.6.2.2.

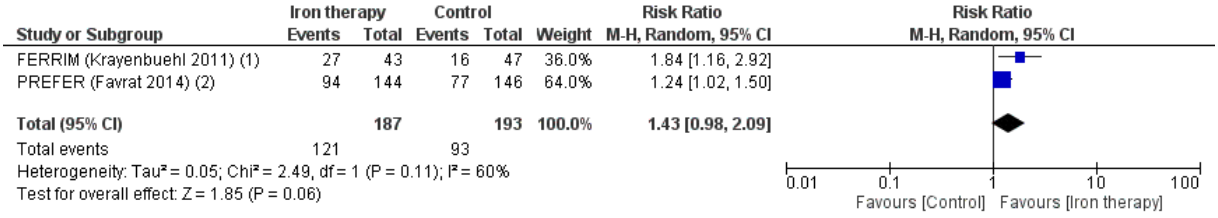


Footnotes
(1) BFI; range from no [0] to maximum imaginable fatigue [10]; median change from baseline, neg. values indicate improvement; 12 weeks; SD estimated from IQR
(2) 22-item Piper Fatigue Scale; range from no fatigue at all [1] to very severe fatigue [10]; mean change from baseline, neg. values indicate improvement; 8 weeks
(3) Global fatigue index - MAF score; range 0-50; change from baseline, negative values indicate improvement; 12 weeks
(4) VAS; Range 1-10, 1 indicates no fatigue at all and 10 very severe fatigue; mean change from baseline, neg. values indicate improvement; 4 weeks

Figure 18 Women with fatigue, fatigue severity

2.3.3.4.2 Fatigue improvement

Two RCTs (FERRIM, PREFER) reported on fatigue improvement. The FERRIM RCT used the Short Performance Inventory to assess the rate of patients with fatigue improvement at 12-week follow-up. The PREFER RCT defined fatigue improvement by the number of patients with a ≥1 point decrease on the Piper Fatigue Scale at eight weeks follow-up. The pooled relative risk for fatigue improvement of iron therapy compared to placebo 1.43 (CI 0.98, 2.09, Figure 8; very low quality of evidence, Table 23).



Footnotes
(1) Fatigue improved assessed with Short performance inventory; 12 weeks
(2) Piper Fatigue Scale: ≥1 point decrease in total was defined as improvement; 8 weeks

Figure 19 Women with fatigue, fatigue improvement

2.3.3.4.3 Adverse events

Adverse events were reported in three RCTs (FERRIM, PREFER, Vaucher 2012). There was no statistically significantly increased risk of adverse events for iron therapy compared to placebo (RR 1.08, 95% CI [0.80, 1.44], Figure 29). Heterogeneity between the RCTs was high (I^2=59%). Excluding FERRIM decreases the I^2=0%; however, it remains unclear why the FERRIM trial reported more adverse events in the control group than in the iron therapy group.

Additional results on adverse events can be found in section 2.3.5.1.1

2.3.3.4.4 Serious adverse events

A total of eight serious adverse events were reported in three RCTs (FERRIM, PREFER, Vaucher 2012). Vaucher 2012 reported four hospitalizations (abdominoplasty, pregnancy, thyroid adenoma and gynaecological surgery) and one severe traffic accident. FERRIM reported one event of appendicitis and one traffic accident, while PREFER reported one event of moderate left thoracic pain. The pooled relative risk for serious adverse events of iron therapy compared to placebo was RR 0.95 (95% CI [0.25, 3.64], Figure 30). Heterogeneity between RCTs was low (I^2=0%). Two RCTs (PREFER, Vaucher 2012) explicitly stated that were no deaths.

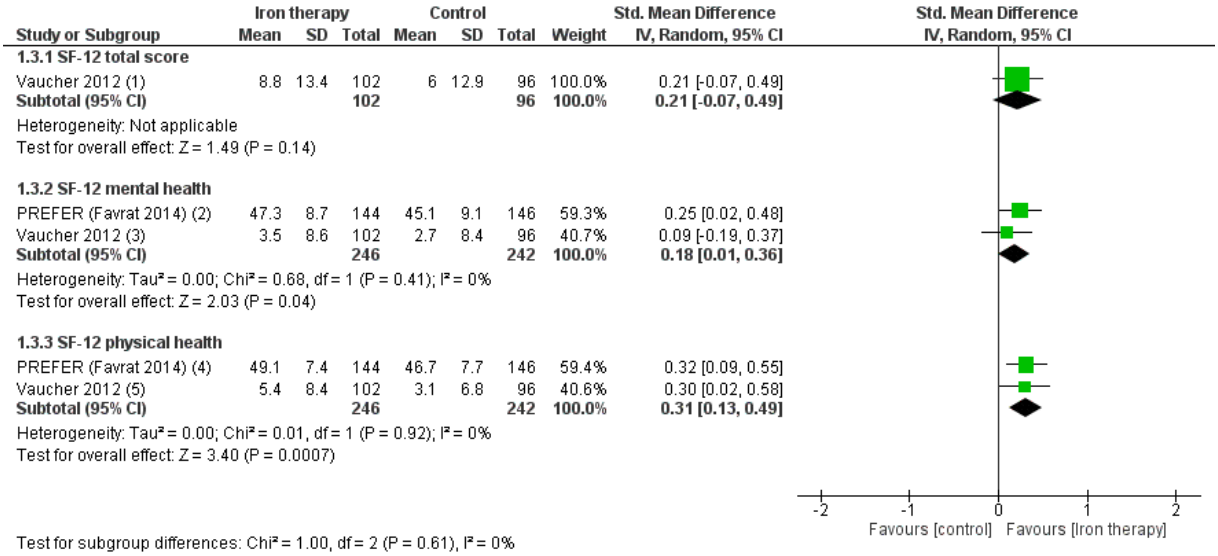
Additional results on adverse events can be found in section 2.3.5.1.2.

2.3.3.5 Important outcomes

2.3.3.5.1 Quality of life

Two RCTs (PREFER, Vaucher 2012) reported on quality of life at 12 weeks (PREFER) and eight weeks (Vaucher 2012) follow-up. Both RCTs used the Short-Form 12 for mental health (SF-12 mental health score, range 0 to 50 [higher score indicating better quality of life]) and physical health (SF-12 physical health score, range 0 to 50 [higher score indicating better quality of life]). Compared to placebo, iron therapy showed a statistically significant increase in mental health scores (SMD 0.18, 95% CI [0.01, 0.36], Figure 9; moderate quality of evidence, Table 23) and physical health scores (SMD 0.31, 95% CI [0.13, 0.49], Figure 9; moderate quality of evidence, Table 23). Heterogeneity between the RCTs was low (I^2=0%) for both SF-12 mental health and physical health scores.

Vaucher 2012 also reported a SF-12 total score (SF-12 total score, range 0 to 100 [higher score indicating better quality of life]) and found a statistically not significant increase in the SF-12 total score in patients with comparing iron therapy compared to control (SMD 0.21, 95% CI [-0.07, 0.49], Figure 9; moderate quality of evidence, Table 23).



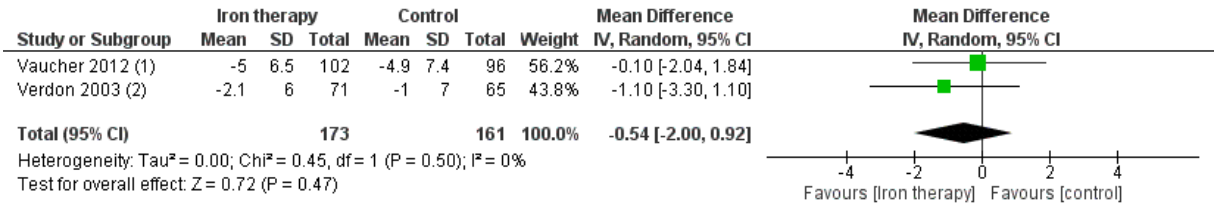
Footnotes

(1) SF-12 total score; range 0-100, higher score indicate better QoL; change from baseline, positive values indicate improvement; 12 weeks
(2) SF-12 mental health score; range n.r., higher score indicate better QoL; 8 weeks
(3) SF-12 mental health score; range 0-50, higher score indicate better QoL; change from baseline, positive values indicate improvement; 12 weeks
(4) SF-12 physical health score; range n.r., higher score indicate better QoL; 8 weeks
(5) SF-12 physical health score; range 0-50, higher score indicate better QoL; change from baseline, positive values indicate improvement; 12 weeks

Figure 20 Women with fatigue, Quality of life

2.3.3.5.2 Depression

Two RCTs (Vaucher 2012, Verdon 2003) reported on depression at 12 weeks (Vaucher 2012) and four weeks (Verdon 2003) follow-up. Both RCTs used the depression subscale of the Current and Past Psychological Survey (CAPPS, range from 0 to 40 [higher score indicating more depressive]). Compared to placebo, iron therapy did not lead to a statistically significant reduction in depression scores (MD -0.54, 95% CI [-2.00, 0.92], Figure 10; low quality of evidence, Table 23). Heterogeneity between the RCTs was low (I^2=0%).



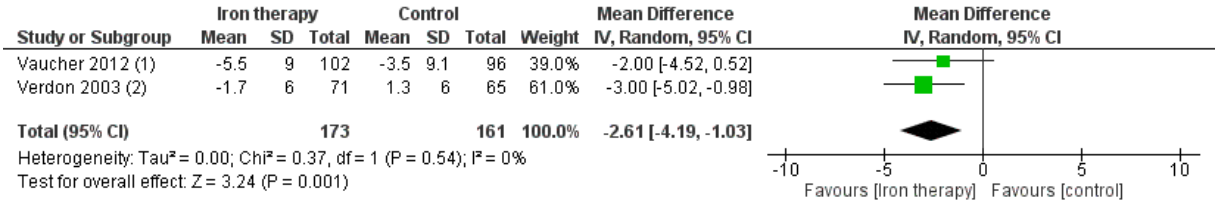
Footnotes

(1) Depression subscale from CAPPS; range 0-40; change from baseline, negative values indicate improvement; 12 weeks
(2) Depression subscale from CAPPS; range 0-40; change from baseline, negative values indicate improvement; 4 weeks

Figure 21 Women with fatigue, Depression

2.3.3.5.3 Anxiety

Two RCTs (Vaucher 2012, Verdon 2003) reported on anxiety at 12 weeks (Vaucher 2012) and four weeks (Verdon 2003) follow-up. Both RCTs used the anxiety subscale of the Current and Past Psychological Survey (CAPPS, range from 0 to 40 [higher score indicating more anxious]). Compared to placebo, iron therapy lead to a statistically significant reduction of anxiety scores (MD -2.61, 95% CI [-4.19, -1.03], Figure 11; low quality of evidence, Table 23). Heterogeneity between the RCTs was low (I^2=0%).



Footnotes
(1) Anxiety subscale from CAPPS; range 0-40; change from baseline, negative values indicate improvement; 12 weeks
(2) Anxiety subscale from CAPPS; range 0-40; change from baseline, negative values indicate improvement; 4 weeks

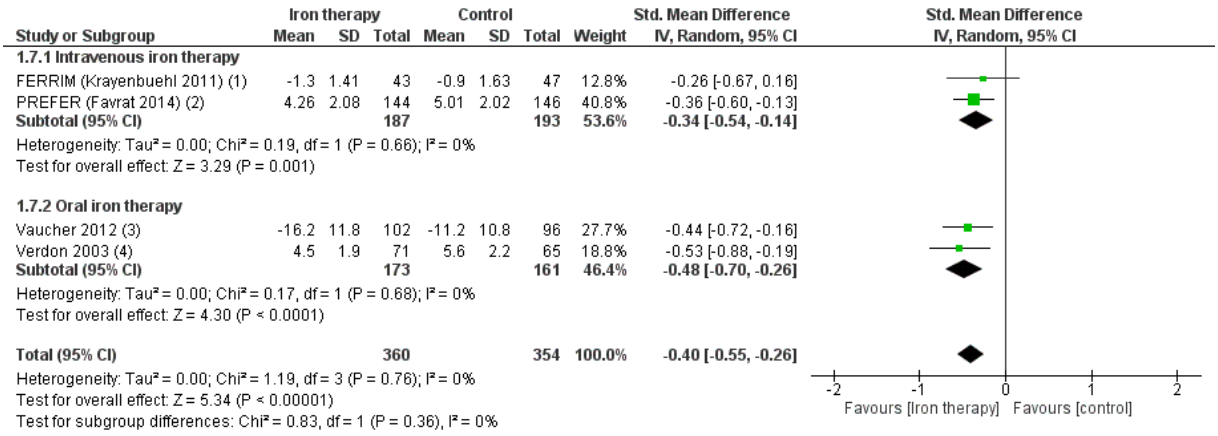
Figure 22 Women with fatigue, anxiety

2.3.3.6 Subgroup analyses

Although the number of included RCTs is very small the pre-specified subgroup analyses are presented here. Due to the limited number of RCTs per subgroup, effect estimates in subgroup need to be interpreted with care and no conclusion should be drawn (As only women were included in the RCTs of interest subgroup analyses for men or children were not applicable).

2.3.3.6.1 Subgroup 1: Oral vs. intravenous therapy with iron

No statistically significant difference in the reduction of fatigue severity was found between trials using intravenous iron and oral iron administration (P = 0.36). Moreover, route of administration has no impact on the association between fatigue severity (standardised differences) and baseline ferritin concentration (for more information on the IPD meta-analysis see the following section 2.3.3.6.2.2).



Footnotes
(1) BFI; range from no [0] to maximum imaginable fatigue [10]; median change from baseline, neg. values indicate improvement; 12 weeks; SD estimated from IQR
(2) 22-item Piper Fatigue Scale; range from no fatigue at all [1] to very severe fatigue [10]; 8 weeks
(3) Global fatigue index - MAF score; range 0-50; change from baseline, negative values indicate improvement; 12 weeks
(4) VAS; Range 1-10, 1 indicates no fatigue at all and 10 very severe fatigue; 4 weeks

Figure 23 Women with fatigue, fatigue severity, subgroup: route of administration

2.3.3.6.2 Subgroup 2: Iron status of study population at recruitment

Fatigue severity was reported by all four trials. First, subgroups were assessed based on aggregated data whereas trials were grouped by their inclusion criteria for ferritin concentration. FERRIM and Vaucher 2012 recruited women with serum ferritin concentrations below 50 µg/l, PREFER below 15 µg/l and Verdon 2003 did not specify an upper ferritin concentration limit, but 85% of the women in

Verdon 2003 were below 50 µg/l at baseline (see section 2.3.3.6.2.1). Second, subgroups were assessed based on IPD (see details in section 2.3.3.6.2.2).

2.3.3.6.2.1 Fatigue severity: Subgroups of trial-specific (aggregated) data

No statistically significant difference in the reduction of fatigue severity was found between trials including women with serum ferritin concentrations below 50 µg/l (FERRIM, Vaucher 2012, Verdon 2003) and one trial (PREFER) which included women with serum ferritin concentration lower than 15 µg/l (P = 0.67).

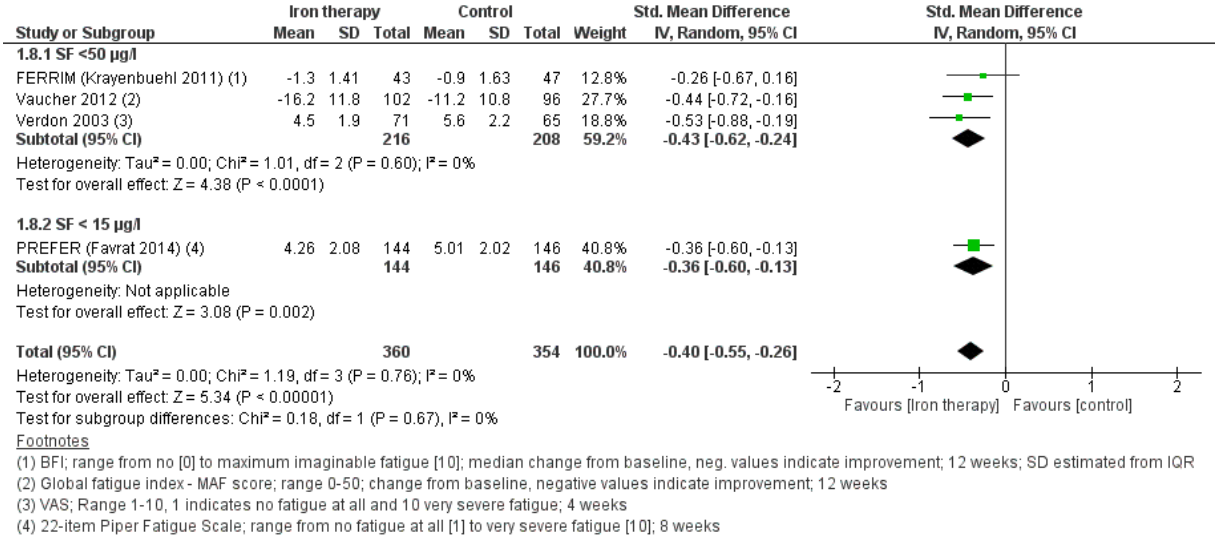


Figure 24 Women with fatigue, fatigue severity, subgroup: iron status of the study population at recruitment

2.3.3.6.2.2 Fatigue severity: Individual patient data meta-analysis

Individual patient data providers and available iron deficiency biomarkers

Anonymized individual patient data was provided by the investigators (Verdon 2003) or the sponsor (Vifor AG: FERRIM and PREFER; Pierre Fabre: Vaucher 2012). At baseline, ferritin concentration was the only iron deficiency biomarker reported by all four trials and was thus considered for IPD analysis. The following biomarkers for iron deficiency were not reported by all trials at baseline: red blood cell count (three trials), reticulocytes (one trial), serum iron (two trials), soluble transferrin receptor (two trials), transferrin (two trials), transferrin saturation (two trials), and total iron-binding capacity (one trial). Haemoglobin (supposed to be normal as it was an inclusion criteria by all trials), haematocrit and mean corpuscular volume were available for all four trials. A complete list of available variables per trial and measured time point is summarized in Appendix 5.4 in Table 34.

Completeness of data for IPD meta-analysis

Measures of baseline and follow-up fatigue severity, and baseline ferritin concentration were completely available for 657 patients out of 718 patients (91.5%). Data was not imputed mainly because the recording of the data-sets of the individual trials differed substantially which did not allow to impute data without introducing uncertainty. Consequently, for the analysis, complete case data

was used. Detailed information on missing data per trial are presented in Table 12. Information on baseline ferritin concentrations, difference of fatigue severity (baseline to last follow-up visit) and the standardised scores of the difference for each study are listed in Table 12. 42.8% (281/657) of the women had a baseline ferritin concentration of <16 µg/l, 32.1% (211/657) between 16 and <30 µg/l, 20.5% (135/657) between 30 and <50 µg/l and 4.6% above ≥50 µg/l.

IPD meta-analysis – Ferritin concentration at baseline

Figure 25 shows the distribution of the standardized difference of fatigue severity (baseline to last follow-up visit) according to the baseline ferritin concentration in the intervention and the control groups of the four trials). When fitting an unadjusted linear regression a very weak association of fatigue improvement by baseline ferritin concentration is suggested (see Figure 40 in Appendix 5.4). The association of ferritin concentrations at baseline and the standardized difference of fatigue severity was further assessed with a multilevel mixed linear regression model taking into account random effects (trial level) and fixed effects (group allocation, ferritin concentration at baseline, follow-up period, route of administration). The estimate for the intervention vs. control group was statistical significant (-0.361, 95% CI [-0.511 to -0.211], P < 0.001, Table 13), but no association between baseline ferritin concentrations and difference of fatigue severity was found (estimate: 0.001, 95% CI [-0.004 to 0.005], P = 0.733, Table 13). In addition, no non-linear relation between baseline ferritin concentrations and difference of fatigue severity was identified. In IPD-sensitivity analyses no association between baseline ferritin concentrations and difference of fatigue severity was found when excluding women with baseline ferritin concentration >100 µg/l (n = 5), when restricting the analysis to parenteral or oral iron therapy, or when using the original fatigue scales by running four linear regressions for each trial separately (for more details see Table 35 in Appendix 5.4).

In two additional analysis, women were categorised into two groups of baseline ferritin concentrations: <16 µg/l (n = 281) vs ≥16 µg/l (n = 376), and <30 µg/l (n = 492) vs ≥30 µg/l (n = 165). As the upper limit for baseline ferritin concentration was defined by the inclusion criteria of the trials (50 µg/l: FERRIM, Vaucher 2012 and Verdon 2003; and 15 µg/l: PREFER), the third additional analysis (ferritin <50 µg/l vs ≥50 µg/l) was considered too explorative (too small number of available measures) and was not conducted. Women with a ferritin concentration <16 µg/l had a slightly greater benefit than women with a ferritin concentration ≥16 µg/l, but the 95% CI was wide, included the null line and was not statistical significant (estimate: -0.104, 95% CI [-0.258 to 0.049], P = 0.182, Table 14). Women with a ferritin concentration <30 µg/l had no benefit when compared to women with a ferritin concentration ≥30 µg/l (estimate: -0.020, 95% CI [-0.154 to 0.194], P = 0.823, Table 15).

IPD meta-analysis – Haemoglobin, haematocrit and mean corpuscular volume at baseline

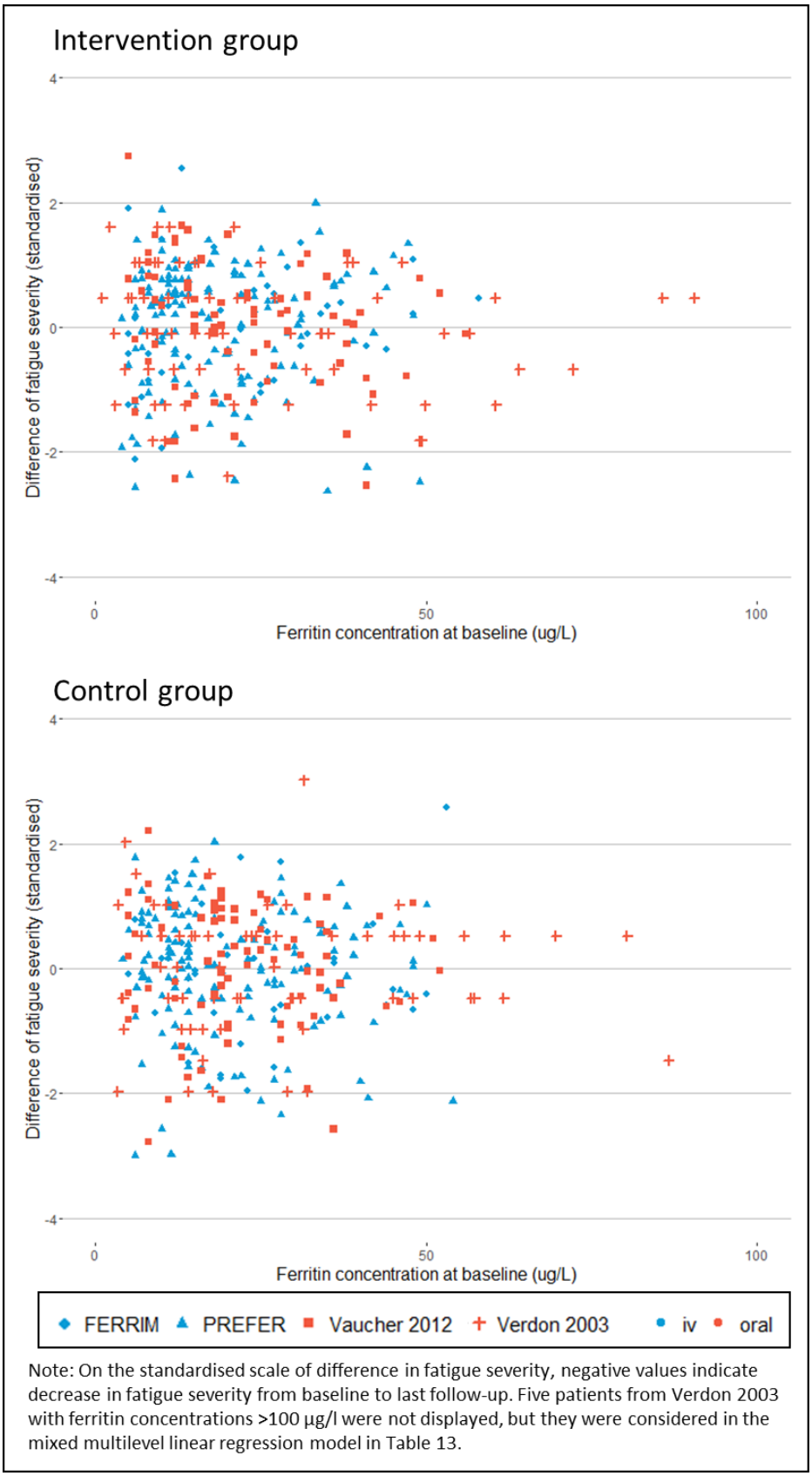
The same multilevel mixed linear regression model was run for the erythrocyte baseline parameters haemoglobin concentration, haematocrit and mean corpuscular volume that were available in all four trials at baseline. Haemoglobin concentration was an inclusion criteria of the trials and was supposed to be normal, and haematocrit and mean corpuscular volume are strongly related to haemoglobin concentration. No association was found between those biomarkers and difference of fatigue severity, the results are presented in Table 36 in Appendix 5.4. Other biomarkers like C-reactive protein, transferrin, transferrin receptor, etc. were not reported by all four trials. Reporting of adverse and serious adverse events was very diverse between trials and missing at all for one trial (Verdon 2003). Therefore, it was concluded that the information gained from the individual patient data will not add

any additional information to what has been reported in the original publications of the other three trials.

Table 12 Standardised scores of the difference of fatigue severity from baseline to latest follow-up visit

Trial name	Time point Complete data*	 Median (Q1, Q3) serum ferritin concentration at baseline**	Instrument Original score (mean difference from baseline±SD)	 Standardised score (mean±SD)***
FERRIM (Krayenbuehl 2011)	12 weeks Intervention group: 37/43 Control group: 38/47	 Intervention group: 22.0 µg/l (10.0, 31.0) Control group: 22.0 µg/l (14.0, 32.0)	Brief Fatigue Inventory Intervention group: -1.50±1.75 Control group: -1.27±1.82	 Intervention group: -0.07±0.99 Control group: 0.06±1.02
PREFER (Favrat 2014)	12 weeks Intervention group: 142/145 Control group: 142/149	 Intervention group: 14.1 µg/l (10.0, 24.0) Control group: 16.0 µg/l (11.4, 28.0)	Global fatigue index – MAF Intervention group: -2.21±2.14 Control group: -1.38±2.03	 Intervention group: -0.20±1.01 Control group: 0.20±0.96
Vaucher 2012	8 weeks Intervention group: 81/102 Control group: 82/96	 Intervention group: 19.0 µg/l (12.0, 31.0) Control group: 20.0 µg/l (14.0, 31.0)	22-item Piper Fatigue Scale Intervention group: -15.86±11.77 Control group: -11.60±10.48	 Intervention group: -0.19±1.04 Control group: 0.19±0.93
Verdon 2003	4 weeks Intervention group: 71/71 Control group: 64/65	 Intervention group: 19.3 µg/l (9.4, 41.7) Control group: 21.7 µg/l (12.9, 38.5)	VAS Intervention group: -1.82±1.75 Control group: 1.05±2.00	 Intervention group: -0.19±0.92 Control group: 0.21±1.05
Total (All four trials)	n.a. Intervention group: 331/361 Control group: 326/357	 Intervention group: 17.3 µg/l (10.0, 29.0) Control group: 19.0 µg/l (12.0, 30.0)	n.a.	 Intervention group: -0.18±0.99 Control group: 0.18±0.97

*Measures of baseline and follow-up fatigue severity, and baseline ferritin concentration were completely reported. **Reported ferritin concentrations might slightly deviate from the study-specific ferritin concentrations reported by the study authors in the original publications because only a complete case scenario (no missing values in fatigue outcome) was required for the IPD. ***Individual fatigue scores were subtracted by the trial mean fatigue score and then divided by its standard deviation. Abbreviations: n.a., not applicable; MAF, Multidimensional Assessment of Fatigue score; SD, standard deviation; VAS, visual analogue scale.



Note: On the standardised scale of difference in fatigue severity, negative values indicate decrease in fatigue severity from baseline to last follow-up. Five patients from Verdon 2003 with ferritin concentrations >100 µg/l were not displayed, but they were considered in the mixed multilevel linear regression model in Table 13.

Figure 25 Distribution of differences of fatigue severity (standardised) from baseline to last follow-up and baseline ferritin concentration

Table 13 Multilevel linear mixed model for difference in fatigue severity (standardised) and ferritin concentration as continuous variable

Parameter	Estimates (95% CI)	P-value
Group (intervention vs. control)	-0.361 (-0.511 to -0.211)	<0.001
Ferritin concentration at baseline (µg/l)	0.001 (-0.004 to 0.005)	0.733
Follow-up in days	-0.002 (-0.005 to 0.002)	0.338
Route of administration (parenteral vs. oral)	0.002 (-0.151 to 0.154)	0.982

Note: see supporting material in Appendix Section 5.4 and sensitivity analyses (Table 35).

Table 14 Multilevel linear mixed model for difference in fatigue severity (standardised) and baseline ferritin concentration categorized into subgroups of concentration of <16 µg/l or ≥16 µg/l.

Parameter	Estimates (95% CI)	P-value
Group (intervention vs. control)	-0.371 (-0.521 to -0.221)	<0.001
Ferritin subgroups:<16 µg/l vs. ≥16 µg/l*	-0.104 (-0.258 to 0.049)	0.182
Follow-up in days	-0.001 (-0.005 to 0.002)	0.390
Route of administration (parenteral vs. oral)	-0.014 (-0.165 to 0.137)	0.855

*upper limit of baseline ferritin defined by exclusion criteria of the trials, see Table 9. The proportion of the women with baseline <16 µg/l was 42.8% vs. 57.2% with ≥16 µg/l.

Table 15 Multilevel linear mixed model for difference in fatigue severity (standardised) and baseline ferritin concentration categorized into subgroups of concentration of <30 µg/l or ≥30 µg/l.

Parameter	Estimates (95% CI)	P-value
Group (intervention vs. control)	-0.362 (-0.512 to -0.212)	<0.001
Ferritin subgroups: <30 µg/l vs. ≥30 µg/l*	-0.020 (-0.154 to 0.194)	0.823
Follow-up in days	-0.002 (-0.005 to 0.002)	0.327
Route of administration (parenteral vs. oral)	-0.001 (-0.152 to 0.151)	0.993

*upper limit of baseline ferritin defined by exclusion criteria of the trials, see Table 9. The proportion of the women with baseline <30 µg/l was 74.9% vs. 25.1% with ≥30 µg/l.

2.3.4 Children with attention-deficit hyperactivity disorder

Characteristics, risk of bias assessment and results for each outcome are shown in the following sections.

2.3.4.1 Overview of included RCTs

One RCT was identified. The reference, analysed outcomes and follow-up time are provided in Table 16 and Table 17.

Table 16 Children with ADHD: Overview of included RCTs, their RCT names and references

Trial name	Reference (Main reference highlighted in colour)
Konofal 2008[57]	Konofal E, Lecendreux M, Deron J, et al. Effects of iron supplementation on attention deficit hyperactivity disorder in children. *Pediatr Neurol.* 2008;38(1):20-26.

Table 17 Children with ADHD: Overview of the included outcomes with analysed follow-up time-points

Outcome / Trial name	ADHD	Clinical Global Impression	Restless legs syndrome Diagnosis	Adverse Events
Konofal 2008	12	12	12	12

The numbers in the fields denote the analysed follow-up period in weeks.

2.3.4.2 Characteristics of the included RCTs

One RCT was identified measuring the effects of iron supplementation in non-anaemic, iron-deficient children with attention-deficit hyperactivity disorder (ADHD). A summary of the RCT characteristics and select baseline characteristics of the patients can be found in Table 18 and Table 19, respectively. The RCT was conducted in France in 2004 and had a 12 week follow-up period. Of the 23 children with ADHD, 18 were randomised to iron therapy and five were randomised to placebo (3:1 randomisation ratio). The patient flow in the intervention group was not clearly reported, baseline characteristics were only reported in 17 children, and the outcome of RLS was reported in 19 children. Industry involvement is unclear, one co-author of the main publication had an industry affiliation.

2.3.4.3 Risk of bias

The risk of selection bias (method of random sequence generation and allocation concealment) was unclear. The risks of performance bias and detection bias were low, while the risk of attrition bias (for both continuous and binary outcome data) was high. Risk of reporting bias was unclear, because no trial protocol was found. A summarised version of the risk of bias assessment is shown in Table 20 and a detailed summary with support of judgment is in Appendix 3.

Table 18 Children with ADHD: Baseline characteristics

Trial name	Intervention Group*	Comparator Group*
Konofal 2008	18 randomised (baseline reported for 17) CPRS total score**: 56.2 ± 11.9 ADHD RS score***: 38.1 ± 6.6 14 males (82%) Age: 5.7 ± 1.2 years Serum ferritin: 29.1 ± 17.6 ng/ml Hb: 12.6 ± 0.8 g/dl	5 randomised CPRS total score**: 79.0 ± 29.0 ADHD RS score***: 35.0 ± 8.0 3 males (60%) Age: 6.4 ± 0.9 years Serum ferritin: 26.2 ± 10.2 ng/ml Hb: 12.8 ± 0.6 g/dl

*data are shown as mean ± standard deviation, unless otherwise specified; **CPRS total score, Conners' Parent Rating Scale total score, range 0-144, higher score indicates more severe ADHD; ***ADHD RS score, Attention-Deficit Hyperactivity Disorder Rating Scale score, range 0-54, higher score indicates more severe ADHD
Abbreviations: ADHD, Attention-Deficit Hyperactivity Disorder; Hb, haemoglobin

Table 19 Children with ADHD: Summary of RCT characteristics and intervention

Trial name	Setting Enrollment period Time-points of FU	Population Key inclusion criteria*	Intervention Compound Dosage regimen	Comparator Compound Dosage regimen
Konofal 2008 France	Child and Adolescent Psychopathology Service of the Hospital Robert Debré (Outpatient) May 2004 – Dec. 2004 (study period) 4, 8 and 12 weeks	Non-anaemic iron-deficient children with ADHD DSM-IV diagnostic criteria for ADHD Serum ferritin <30 ng/ml Normal haemoglobin levels IQ ≥80 on the Wechsler Intelligence Scale Age: 5-8 years No relevant psychiatric comorbidities	Ferrous sulfate (Tardyferon®, Robapharm) Oral 80 mg capsules once daily for 12 weeks	Placebo tablets (identical even when sliced) Oral capsules once daily for 12 weeks

*see Appendix 5.2 for more details on inclusion and exclusion criteria
Abbreviations: ADHD, attention-deficit hyperactivity disorder, FU, Follow-up

Table 20 Risk of bias, Children with ADHD

Trial name	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete continuous outcome data (attrition bias)	Incomplete binary data (attrition bias)	Selective reporting (reporting bias)
Konofal 2008	Unclear	Unclear	Low	Low	High	High	Unclear

2.3.4.4 Critical outcomes

2.3.4.4.1 ADHD severity

Konofal 2008 assessed ADHD severity with the Conner's Parent Rating scale (CPRS, range n.r., higher score indicates more severe symptoms of ADHD). Compared to control, iron therapy had no statistically significant effect on ADHD symptom severity reduction (MD -3.80, 95% CI [-24.86, 17.26], Figure 26; very low quality of evidence, Table 24). Konofal 2008 reported five additional measures for ADHD symptoms which were summarized in Table 21. Information on the quality of evidence for these outcomes can be found in Table 24.

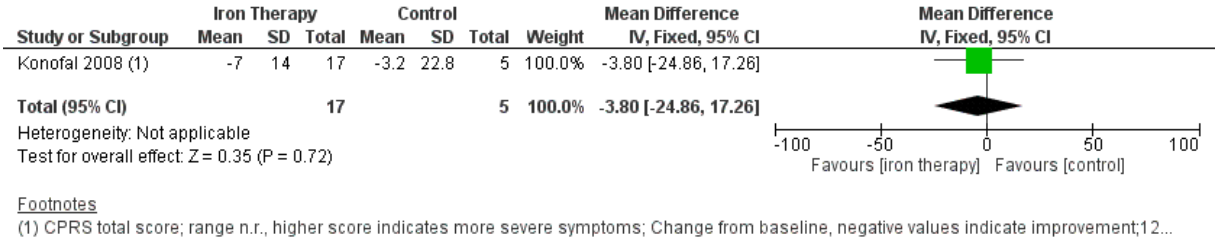


Figure 26 Children with ADHD, Symptoms of ADHD severity

Table 21 Children with ADHD, symptoms of ADHD severity

Instrument	Range, direction	Iron Therapy (n=17) Mean change from baseline (SD)	Control (n=5) Mean change from baseline (SD)	Statistical significance of mean difference*
Conner's Parent Rating Scale (CPRS) total score	Range n.r., higher score indicates more severe symptoms of ADHD	-7.0 (14.0)	-3.2 (22.8)	not significantly different
ADHD index	Range and direction n.r.	-1.8 (3.8)	-0.2 (7.2)	not significantly different
Conner's Teacher Rating Scale (CTRS) total score	Range n.r., higher score indicates more severe symptoms of ADHD	-5.3 (11.2)	2.0 (3.4)	not significantly different
ADHD Rating Scale (total score)	Range n.r., higher score indicates worse symptoms of ADHD	-10.2 (14.0)	-3.0 (5.7)	not significantly different
Inattentive subscore of the ADHD Rating Scale	Range n.r., higher score indicates worse symptoms of ADHD	-4.4 (7.0)	-0.8 (2.5)	not significantly different
Hyperactive/Impulsive subscore of the ADHD Rating Scale	Range n.r., higher score indicates worse symptoms of ADHD	-5.8 (7.5)	-2.2 (3.7)	not significantly different

*based on reported 95% confidence intervals
Abbreviations: ADHD, Attention-Deficit Hyperactivity Disorder; n, number of participants; n.r., not reported; SD, standard deviation

2.3.4.4.2 Adverse events

Konofal 2008 reported the number of participants with adverse events at 12 weeks follow-up. There was no statistically significant difference in adverse events between iron therapy and placebo (RR 0.42, 95% CI [0.09, 1.85], Figure 29).

Additional results on adverse events can be found in section 2.3.5.1.1

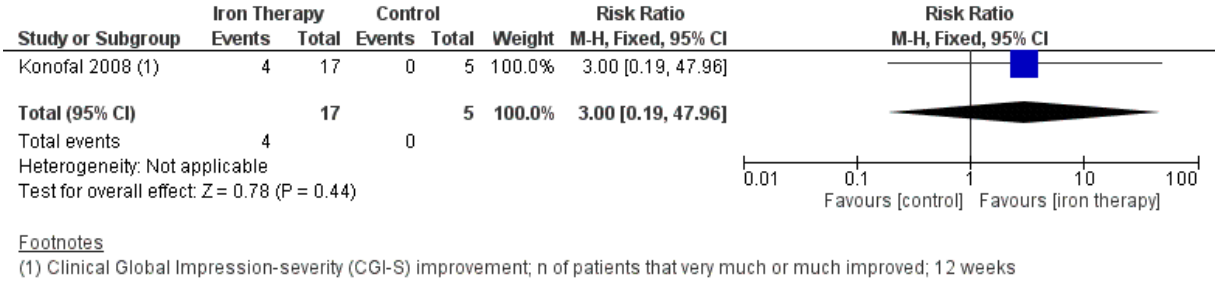
2.3.4.4.3 Serious adverse events

Information on serious adverse events was lacking.

2.3.4.5 Important outcomes

2.3.4.5.1 Clinical global impression improvement

Konofal 2008 dichotomized the Clinical Global Impression-Severity (CGI-S) to assess overall improvement at 12 weeks follow-up. Four out of 17 children with iron therapy and no out of 5 children in the control group showed an improvement in the clinical global impression (RR 3.00, 95% CI [0.19, 47.96], Figure 27; very low quality of evidence, Table 24).

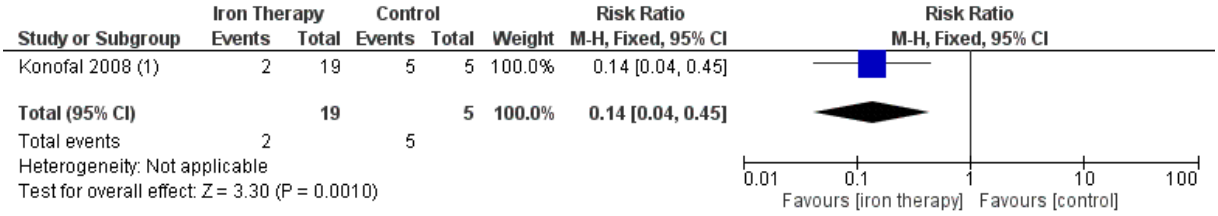


Footnotes
(1) Clinical Global Impression-severity (CGI-S) improvement; n of patients that very much or much improved; 12 weeks

Figure 27 Children with ADHD, global impression improvement

2.3.4.5.2 Restless legs syndrome diagnosis

Presence of RLS in children at trial termination at 12 weeks was diagnosed with the International Restless Legs Syndrome Study Group criteria. Two out of 19 children with iron therapy and all 5 children in the control group were diagnosed with RLS (RR 0.14, 95% CI [0.04, 0.45], Figure 27; low quality of evidence, Table 24).



Footnotes
(1) RLS diagnosis according to the International Restless Legs Syndrome Study Group criteria specific for children; 12 weeks

Figure 28 Children with ADHD, restless legs syndrome diagnosis

2.3.4.6 Subgroup analyses

With one RCT on children with ADHD no subgroup analyses were performed.

2.3.5 Safety outcomes, all populations

2.3.5.1 Critical outcomes

2.3.5.1.1 Adverse events (in all patient populations)

Five RCTs (Allen 2011, FERRIM, Grote 2009, Konofal 2008, Vaucher 2012) reported on the number of patients with adverse events and two RCTs (PREFER, Trenkwalder 2017) reported on the number of patients with adverse reaction or treatment emergent adverse events. All events were reported within a range of four to 52 weeks follow-up. Patients with iron compared to control were statistically not significantly more likely to experience adverse events (RR 1.12, 95% CI [0.88, 1.41], Figure 29; low quality of evidence, Table 25). Heterogeneity between the RCTs was low (I^2=34%). The test for subgroup differences identified no statistically significant effects (P = 0.28) between the three symptomatic IDNA populations.

Also, within the three different study populations, no statistically significant differences between iron therapy and placebo were found.

2.3.5.1.1.1 Adverse events in the restless legs syndrome population

Three RCTs (Allen 2011, Grote 2009, Trenkwalder 2017) reported the number of patients with adverse events. Patients receiving iron therapy compared to control/placebo arm were statistically significantly more likely to experience adverse events (RR 1.37, 95% CI [0.88, 2.13], Figure 29). Heterogeneity between the RCTs was low (I^2=0%).

Davis 2000 and Earley 2009 only reported the number of adverse events and side effects, respectively. Davis 2000 and Earley 2009 were therefore not pooled with the other RCTs reporting the number of patient with adverse events. Davis 2000 reported 12 adverse events at 14 weeks follow-up in the iron therapy group in a total of 14 randomised patients and zero adverse events in the placebo group in a total of 14 randomised patients. Earley 2009 reported on the day of infusion 13 side effects in the iron therapy group in a total of 11 randomised patients and two side effects in the placebo group in a total of 7 randomised patients. All side effects resolved within minutes or hours after infusion. Earley 2009 described no adverse effects in the iron therapy group and placebo group at two weeks follow-up.

2.3.5.1.1.2 Adverse events in women with fatigue

Three RCTs (FERRIM, PREFER, Vaucher 2012) reported adverse events. The relative risk for adverse events of iron therapy compared to placebo was statistically not significantly increased (RR 1.08, 95% CI [0.80, 1.44], Figure 29). Heterogeneity between the RCTs was high (I^2=59%). Excluding the FERRIM trial decreased the I^2=0%; however, and this reported more adverse events in the control than in the iron therapy group.

2.3.5.1.1.3 Adverse events in children with ADHD

Konofal 2008 reported the number of children with adverse events at 12 weeks follow-up. Three out of 18 children with iron therapy and two of 5 children in the control group had adverse events (RR 0.42, 95% CI [0.09, 1.85], Figure 29).

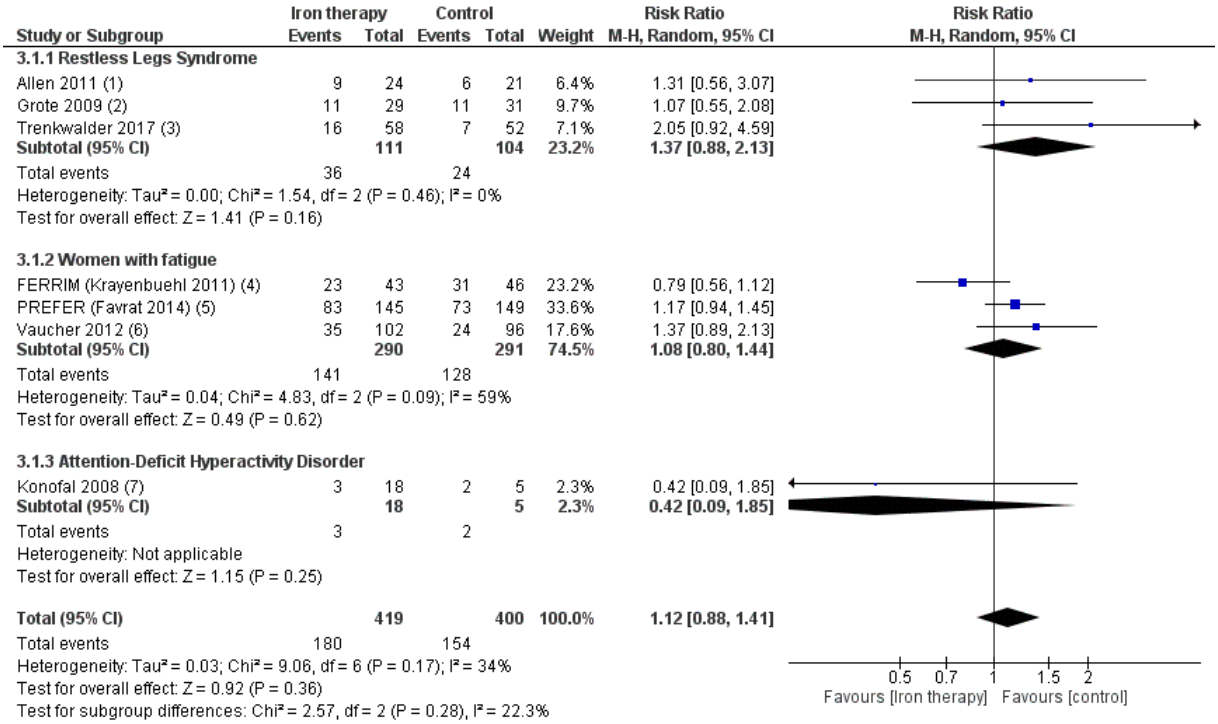


Figure 29 Adverse events, all populations

2.3.5.1.2 Serious adverse events (in all patient populations)

Seven RCTs (Allen 2011, Cho 2016, Davis 2000, FERRIM, PREFER, Trenkwalder 2017, Vaucher 2012) reported the number of patients with a serious adverse event with a range of 6 weeks to end of study (> 12 months) follow-up. Six of 418 patients in the intervention and five out of 410 patients in the control group experienced a serious adverse event. The relative risk of a serious adverse event of iron therapy compared to placebo was statistically not significantly increased (RR 1.27, 95% CI [0.40, 40.3], Figure 30; very low quality of evidence, Table 25). Heterogeneity between the RCTs was low (I^2=0%) (Interaction test for subgroup differences between the three patient populations P = 0.41).

Overall, ten serious adverse events were specified; thereof, Vaucher 2012 reported four hospitalizations (abdominoplasty, pregnancy, thyroid adenoma and gynaecological surgery) and one severe traffic accident. FERRIM reported one event of appendicitis and one traffic accident. Davis 2000 reported one event of vertebral fracture. PREFER reported one event of moderate left thoracic pain. Trenkwalder 2017 did not specify the reported serious adverse event.

Mortality was explicitly reported in five RCTs (Earley 2009, PREFER, Trenkwalder 2017, Vaucher 2012, Wang 2009) with no deaths reported.

2.3.5.1.2.1 Serious adverse events in restless legs syndrome population

Serious adverse events were reported in four RCTs (Allen 2011, Cho 2016, Davis 2000, Trenkwalder 2017), one vertebral fracture (Davis 2000) and one unspecified serious adverse event (Trenkwalder 2017) and the remaining two trials reported no serious adverse events. The relative risk for serious adverse events of iron therapy compared to control was statistically not significantly increased (RR 2.85, 95% CI [0.31, 26.38], Figure 30). Heterogeneity between the RCTs was low (I^2=0%). Mortality was explicitly mentioned in three RCTs (Earley 2009, Trenkwalder 2017, Wang 2009) with no deaths reported.

2.3.5.1.2.2 Serious adverse events in women with fatigue

Eight serious adverse events were reported in three RCTs (FERRIM, PREFER, Vaucher 2012). Vaucher 2012 reported four hospitalizations (abdominoplasty, pregnancy, thyroid adenoma and gynaecological surgery) and one severe traffic accident. FERRIM reported one event of appendicitis and one traffic accident, while PREFER reported one event of moderate left thoracic pain. The relative risk for serious adverse events of iron therapy compared to placebo was statistically not significantly reduced (RR 0.95, 95% CI [0.25, 3.64], Figure 30). Heterogeneity between the RCTs was low (I^2=0%). Two RCTs (PREFER, Vaucher 2012) explicitly reported on morality. There were no deaths reported.

2.3.5.1.2.3 Serious adverse events in children with ADHD

Information on serious adverse events was lacking.

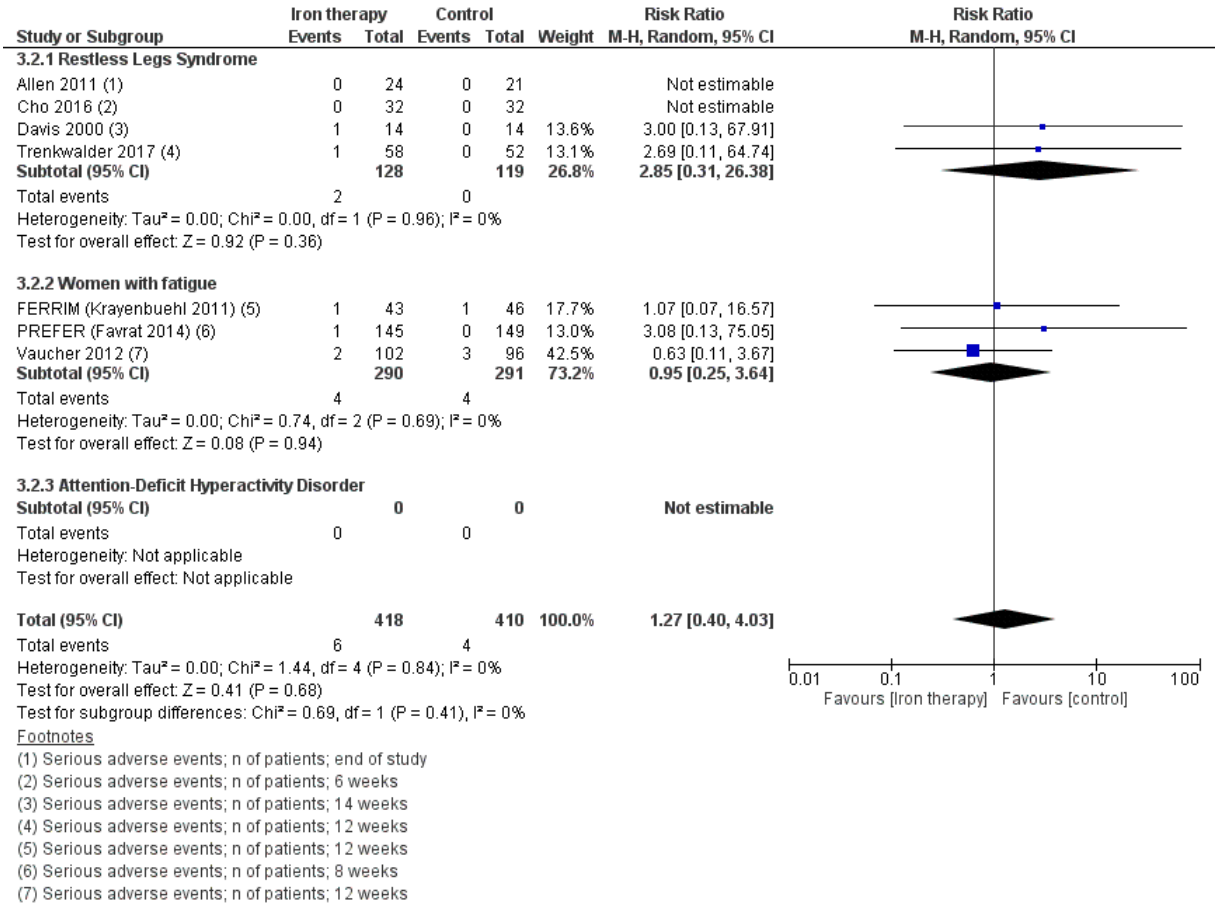


Figure 30 Serious adverse events, all populations

2.3.5.2 Subgroup analyses

Due to the small number of safety outcome events no subgroup analyses were done.

2.4 Summary of findings

2.4.1 Adults with restless legs syndrome

Eight RCTs compared iron therapy to control in adults with RLS. Iron therapy compared to control showed a statistically significant reduction of RLS symptom severity (MD -4.23, 95% CI [-6.11, -2.34]) critical outcome, low quality of evidence, from seven RCTs). A potential "placebo effect" cannot be excluded in six out of seven trials reporting on RLS symptom severity. Moreover, iron therapy compared to control showed a statistically significant improvement in RLS treatment response (RR 1.61, 95% CI [1.13, 2.30], critical outcome, very low quality of evidence, from six RCTs), and of quality of life (SMD 0.51, 95% CI [0.15, 0.87], important outcome, very low quality of evidence, from three RCTs). No statistically significant difference between iron therapy and control was found for sleep quality (critical outcome, very low quality of evidence, from four RCTs), sleepiness (critical outcome, very low quality of evidence, from one RCT), and improvements in depression or anxiety scores (important outcome, very low quality of evidence, each from one trial). Global impression rating with iron therapy compared to control was statistically not significantly different (important outcome, very low quality of evidence, from two RCTs). In contrast, iron therapy compared to control showed a statistically significant increase in the number of patients with improved global impression (important outcome, very low qualities of evidence, from two RCTs).

The relative risk of adverse events of iron therapy compared to control was increased in three RCTs in patients with RLS, but this effect was not statistically significant. The quality of evidence assessment was conducted across all study populations (seven RCTs) and a statistically non-significant increase in adverse events was found (critical outcome, low quality of evidence). A statistically not-significant increase in serious adverse events of iron therapy compared to control was found across four RCTs. For the quality of evidence assessment, the quality of evidence was assessed across all study populations (seven RCTs) and a statistically non-significant increase in serious adverse events was found in patients with iron therapy compared to control (critical outcome, very low quality of evidence).

The overall quality of evidence was judged to be very low because of the very low quality of evidence for the critical outcome of serious adverse events. Additional details are reported in the summary of findings, Table 22.

Table 22 Adults with RLS: Summary of findings (GRADE)

Restless legs syndrome compared to control for iron deficiency without anaemia

Outcomes	№ of participants (RCTs)	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects*	
				Control group risk (only dichotomous outcomes)	Effect estimate (continuous outcomes) and risk difference (dichotomous outcomes) in adults with RLS
Restless legs syndrome Symptom Severity	343 (7 RCTs)	⊕⊕◯◯ LOW [a,b]	-	-	MD **4.23 lower** (6.11 lower to 2.34 lower)
Restless legs syndrome Improvement	325 (6 RCTs)	⊕◯◯◯ VERY LOW [c,d,e]	**RR 1.61** (1.13 to 2.30)	271 per 1'000	**165 more per 1'000** (35 more to 352 more)
Sleep	151 (4 RCTs)	⊕◯◯◯ VERY LOW [f,g,h,i]	-	-	SMD **0.02 lower** (0.46 lower to 0.42 higher)
Quality of life	128 (3 RCTs)	⊕◯◯◯ VERY LOW [i,j,k]	-	-	SMD **0.51 higher** (0.15 higher to 0.87 higher)
Global Impression Rating	128 (2 RCTs)	⊕◯◯◯ VERY LOW [i,l,m,n]	-	-	SMD **0.59 SD lower** (1.25 lower to 0.07 higher)
Global Impression Improvement - Combined Clinical Global Impression and Patient Global Impressions of Improvement	110 (1 RCT)	⊕◯◯◯ VERY LOW [e,o,p]	**RR 2.35** (1.07 to 5.13)	137 per 1'000	**185 more per 1'000** (10 more to 567 more)
Global Impression Improvement - Clinical Global Inventory of Change (CGI-1)	43 (1 RCT)	⊕◯◯◯ VERY LOW [e,q,r]	**RR 3.17** (1.04 to 9.64)	158 per 1'000	**343 more per 1'000** (6 more to 1,364 more)

Restless legs syndrome compared to control for iron deficiency without anaemia

Outcomes	№ of participants (RCTs)	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects*	
				Control group risk (only dichotomous outcomes)	Effect estimate (continuous outcomes) and risk difference (dichotomous outcomes) in adults with RLS
Global Impression Improvement - Patient Global Rating of Change (PGI-1)	43 (1 RCT)	⨁◯◯◯ VERY LOW [e,q,r]	**RR 4.35** (1.09 to 17.33)	105 per 1'000	**353 more per 1'000** (9 more to 1,719 more)
Depression	23 (1 RCT)	⨁◯◯◯ VERY LOW [i,s]	-	-	MD **2.28 SD lower** (8.33 lower to 2.73 higher)
Fatigue	43 (1 RCT)	⨁◯◯◯ VERY LOW [i,p,q]	-	-	MD **1.8 SD higher** (4.71 lower to 8.31 higher)
Sleepiness	23 (1 RCT)	⨁◯◯◯ VERY LOW [i,s]	-	-	MD **0.2 SD lower** (2.66 lower to 2.26 higher)
Adverse events [t]	819 (7 RCTs)	⨁⨁◯◯ LOW [u,v,w]	**RR 1.12** (0.88 to 1.41)	385 per 1'000	**46 more per 1'000** (46 fewer to 158 more)
Serious adverse events [t]	828 (7 RCTs)	⨁◯◯◯ VERY LOW [x,y]	**RR 1.09** (0.35 to 3.37)	12 per 1'000	**1 more per 1'000** (8 fewer to 29 more)

*For dichotomous outcomes, the risk in the intervention group (and its 95% confidence interval) is based on the control group risk and the relative effect of the intervention (and its 95% CI). For continuous outcomes, the severity in the control group was not estimated.

CI: Confidence interval; MD: Mean difference; RR: Risk ratio; RLS: restless legs syndrome; SMD: Standardised mean difference

Restless legs syndrome compared to control for iron deficiency without anaemia

Outcomes	№ of participants (RCTs)	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects*	
				Control group risk (only dichotomous outcomes)	Effect estimate (continuous outcomes) and risk difference (dichotomous outcomes) in adults with RLS

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

a. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 4 RCTs and selection bias (allocation concealment) was unclear in 3 RCTs; risk of performance bias was high in 1 RCT; risk of detection bias was unclear in 4 RCTs; risk of attrition bias was unclear in 2 RCTs and high in 3 RCTs and risk of selective reporting was unclear in 4 RCTs and high in 2 RCTs.
b. Indirectness was serious because iron deficiency status of RCT populations was unclear: in 3 of 7 RCTs iron deficiency was not part of inclusion/exclusion criteria and in 2 of 7 RCTs the included population had a low-normal serum ferritin concentrations (mixed population).
c. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 4 RCTs and selection bias (allocation concealment) was unclear in 2 RCTs; risk of performance bias was high in 1 RCT; risk of detection bias was unclear in 4 RCTs; risk of attrition bias was unclear in 2 RCTs and high in 4 RCTs and risk of selective reporting was unclear in 3 RCTs and high in 2 RCTs.
d. Indirectness was serious because iron deficiency status of study populations was unclear: in 3 of 6 RCTs iron deficiency was not part of inclusion/exclusion criteria and in 1 of 6 RCTs the included population had a low-normal serum ferritin concentrations (mixed population).
e. Imprecision was serious because the total number of events was <300.
f. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 3 RCTs and selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was high in 1 RCT; risk of detection bias was unclear in 4 RCTs; risk of attrition bias was unclear in 1 RCT and high in 3 RCTs and risk of selective reporting was unclear in 4 RCTs.
g. It was not downgraded for inconsistency because heterogeneity was explained during sensitivity analysis where RCTs with a non-placebo comparator were excluded.
h. Indirectness was serious because iron deficiency status of study populations was unclear: in 3 of 4 RCTs iron deficiency was not part of inclusion/exclusion criteria.
i. Imprecision was serious because the total sample size was below the optimal information size (OIS).
j. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 2 RCTs; risk of detection bias was unclear in 3 RCTs; risk of attrition bias was unclear in 1 RCT and high in 2 RCTs and risk of selective reporting was unclear in 3 RCTs.
k. Indirectness was serious because iron deficiency status of study populations was unclear: in 3 of 3 RCTs iron deficiency was not part of inclusion/exclusion criteria.

l. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 1 RCT and selection bias (allocation concealment) was unclear in 1 RCT; risk of attrition bias was unclear in 1 RCT and high in 1 RCT and risk of selective reporting was high in 2 RCTs.
m. It was not downgraded for inconsistency because heterogeneity was low-moderate and confidence intervals were widely overlapping.
n. Indirectness was serious because iron deficiency status of study populations was unclear: in 1 of 2 RCTs iron deficiency was not part of inclusion/exclusion criteria and in 1 of 2 RCTs the included population had a low-normal serum ferritin concentrations (mixed population).
o. The RCT limitation was serious because risk of attrition bias was high in 1 RCT and risk of selective reporting was high in 1 RCT.
p. Indirectness was serious because iron deficiency status of study populations was unclear: in 1 of 1 RCT the included population had a low-normal serum ferritin concentrations (mixed population).
q. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 1 RCT; risk of detection bias was unclear in 1 RCT; risk of attrition bias was high in 1 RCT and risk of selective reporting was unclear in 1 RCT.
r. Indirectness was serious because iron deficiency status of study populations was unclear: in 1 of 1 RCT iron deficiency was not part of inclusion/exclusion criteria.
s. The RCT limitation was very serious because selection bias (random sequence generation) was unclear in 1 RCT and selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was high in 1 RCT; risk of detection bias was unclear in 1 RCT; risk of attrition bias was high in 1 RCT and risk of selective reporting was unclear in 1 RCT.
t. Adverse events and serious adverse events were pooled over all populations.
u. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 4 RCTs and selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was unclear in 1 RCT; risk of detection bias was unclear in 4 RCTs; risk of attrition bias was unclear in 2 RCTs and high in 4 RCTs and risk of selective reporting was unclear in 3 RCTs and high in 1 RCT.
v. It was not downgraded for inconsistency because heterogeneity was low-moderate and confidence intervals were widely overlapping.
w. Imprecision was serious because the 95% CI of the effect estimate is sufficiently wide to include both no effect and appreciable benefit (relative risk increase greater than 25%) in favour of no iron therapy.
x. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 3 RCTs; risk of performance bias was unclear in 1 RCT; risk of detection bias was unclear in 5 RCTs; risk of attrition bias was unclear in 3 RCTs and high in 3 RCTs and risk of selective reporting was unclear in 4 RCTs and high in 1 RCT.
y. Imprecision was serious because the 95% CI of the effect estimate is sufficiently wide to include both appreciable harm or benefit (relative risk increase or decrease greater than 25%) in favour of no iron therapy and because the total number of events was <300.

2.4.2 Women with fatigue

Among the included four RCTs, 42.8% (281/657) of the women had a baseline ferritin concentration of <16 µg/l, 32.1% (211/657) between 16 and <30 µg/l, 20.5% (135/657) between 30 and <50 µg/l and 4.6% above ≥50 µg/l. Four RCTs reported a statistically significant reduction of fatigue severity of iron therapy compared to control (SMD -0.41, 95% CI [-0.56, -0.26], critical outcome, moderate quality of evidence). A potential "placebo effect" cannot be excluded in the trials reporting on fatigue severity. Two RCTs reported on fatigue improvement, iron therapy compared to control showed a statistically not significant increase in fatigue improvement (critical outcome, very low quality of evidence). Two RCTs reported on quality of life and showed a statistically significant improvement in both, mental and physical health scores of iron therapy compared to control (Mental: SMD 0.18, 95% CI [0.01, 0.36] and physical: SMD 0.31, 95% CI [0.13, 0.49], important outcome, moderate qualities of evidence). One RCT showed no statistically significant improvement in quality of life total scores from iron therapy (important outcome, moderate quality of evidence) and in two RCTs, iron therapy compared to control did not statistically significant decrease depression scores (important outcome, low quality of evidence). In two RCTs, iron therapy compared to control showed a statistically significant improvement in anxiety scores (MD -2.61, 95% CI [-4.19, -1.03], important outcome, low quality of evidence).

For patients with fatigue and treated with iron therapy, a statistically not significant increase of the risk for adverse events was seen in three RCTs when compared to control. For the quality of evidence assessment, the quality of evidence was assessed across all study populations (seven RCTs) and a statistically non-significant increase in adverse events was found when comparing iron therapy with control (critical outcome, low quality of evidence). For patients with fatigue and iron therapy, a statistically not significant decrease in serious adverse events was found in RCTs trials when compared to control. For the quality of evidence assessment, the quality of evidence was assessed across all study populations (seven RCTs) and a statistically not significant increase in serious adverse events was found in patients treated with iron therapy compared to control (critical outcome, very low quality of evidence).

The overall quality of evidence was judged to be very low because of the very low quality of evidence for the critical outcome of serious adverse events. Additional details are reported in the summary of findings, Table 23.

Table 23 Women with fatigue: Summary of findings (GRADE)

Women with fatigue compared to placebo for iron deficiency without anaemia

Outcomes	№ of participants (RCTs)	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects*	
				Control group risk (only dichotomous outcomes)	Effect estimate (continuous outcomes) and risk difference (dichotomous outcomes) in women with fatigue
Fatigue severity	714 (4 RCTs)	⨁⨁⨁◯ MODERATE [a]	-	-	SMD **0.41 lower** (0.56 lower to 0.26 lower)
Fatigue improvement	380 (2 RCTs)	⨁◯◯◯ VERY LOW [b,c,d]	**RR 1.43** (0.98 to 2.09)	482 per 1'000	**207 more per 1'000** (10 fewer to 525 more)
Quality of life - SF-12 total score	198 (1 RCT)	⨁⨁⨁◯ MODERATE [e,f]	-	-	SMD **0.21 higher** (0.07 lower to 0.49 higher)
Quality of life - SF-12 mental health	488 (2 RCTs)	⨁⨁⨁◯ MODERATE [g]	-	-	SMD **0.18 higher** (0.01 higher to 0.36 higher)
Quality of life - SF-12 physical health	488 (2 RCTs)	⨁⨁⨁◯ MODERATE [g]	-	-	SMD **0.31 higher** (0.13 higher to 0.49 higher)
Depression	334 (2 RCTs)	⨁⨁◯◯ LOW [f,h,i]	-	-	MD **0.54 lower** (2 lower to 0.92 higher)
Anxiety	334 (2 RCTs)	⨁⨁◯◯ LOW [f,h,i]	-	-	MD **2.61 lower** (4.19 lower to 1.03 lower)
Adverse events [j]	819 (7 RCTs)	⨁⨁◯◯ LOW [l,m,n]	**RR 1.12** (0.88 to 1.41)	385 per 1'000	**46 more per 1'000** (46 fewer to 158 more)

Women with fatigue compared to placebo for iron deficiency without anaemia

Outcomes	№ of participants (RCTs)	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects*	
				Control group risk (only dichotomous outcomes)	Effect estimate (continuous outcomes) and risk difference (dichotomous outcomes) in women with fatigue
Serious adverse events [j]	828 (7 RCTs)	⨁◯◯◯ VERY LOW [o,p]	**RR 1.09** (0.35 to 3.37)	12 per 1,000	**1 more per 1'000** (8 fewer to 29 more)

*For dichotomous outcomes, the risk in the intervention group (and its 95% confidence interval) is based on the control group risk and the relative effect of the intervention (and its 95% CI). For continuous outcomes, the severity in the control group was not estimated.

CI: Confidence interval; RR: Risk ratio; MD: Mean difference; SMD: Standardised mean difference

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations
a. The RCT limitation was serious because selection bias was unclear in 2 RCTs; risk of performance bias was unclear in 1 RCT; risk of detection bias was unclear in two RCTs; risk of attrition bias was unclear in 1 RCT and risk of selective reporting was unclear in 2 RCTs.
b. The RCT limitation was serious because of risk of selection bias was unclear in 1 RCT; risk of performance bias was unclear in 1 RCT; and risk of attrition bias was unclear in 1 RCT.
c. Inconsistency was serious because heterogeneity was high.
d. Imprecision was serious because the 95% CI of the effect estimate is sufficiently wide to include no effect and an appreciable benefit (relative risk increase greater than 25%); in addition, the total sample size did appear lower than the optimal information size (OIS).
e. The RCT limitation was considered as not serious because only risk of reporting bias was unclear.
f. Imprecision was serious because the total sample size was below the optimal information size (OIS).
g. The RCT limitation was serious because risk of performance and detection bias was unclear in 1 RCT; risk of attrition bias was unclear in 1 RCT; and risk of reporting bias was unclear in 1 RCT.
h. The RCT limitation was not serious because only risk of selection bias was unclear in 1 RCT and risk of selective reporting was unclear in 2 RCTs.

i. Indirectness was serious because the instruments that were used to measure depression or anxiety were considered not to be validated.
j. Adverse events and serious adverse events were pooled over all populations.
l. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 4 RCTs and selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was unclear in 1 RCT; risk of detection bias was unclear in 4 RCTs; risk of attrition bias was unclear in 2 RCTs and high in 4 RCTs and risk of selective reporting was unclear in 3 RCTs and high in 1 RCT.
m. It was not downgraded for inconsistency because heterogeneity was low-moderate and confidence intervals were widely overlapping.
n. Imprecision was serious because the 95% CI of the effect estimate is sufficiently wide to include both no effect and appreciable benefit (relative risk increase greater than 25%) in favour of no iron therapy.
o. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 3 RCTs; risk of performance bias was unclear in 1 RCT; risk of detection bias was unclear in 5 RCTs; risk of attrition bias was unclear in 3 RCTs and high in 3 RCTs and risk of selective reporting was unclear in 4 RCTs and high in 1 RCT.
p. Imprecision was serious because the 95% CI of the effect estimate is sufficiently wide to include both appreciable harm or benefit (relative risk increase or decrease greater than 25%) in favour of no iron therapy and because the total number of events was <300.

2.4.3 Children with ADHD

One RCT comparing iron therapy to placebo in children with ADHD showed a statistically non-significant reduction in ADHD severity (critical outcome, very low quality of evidence), a statistically non-significant improvement of the clinical global impression (important outcome, very low quality of evidence) and a statistically significant decrease in the diagnosis of RLS after 12 weeks of therapy (RR 0.14, 95% CI [0.04, 0.45], important outcome, low quality of evidence).

In this trial children with ADHD and treated with iron were slightly less likely to experience adverse events when compared with control (statistically not significant). For the quality of evidence assessment, the quality of evidence was assessed across all study populations in this (seven RCTs) and a statistically non-significant increase of adverse events was found when comparing iron therapy to control (critical outcome, low quality of evidence). No serious adverse events were reported.

The overall quality of evidence was judged to be very low because of the very low quality of evidence for the critical outcome of serious adverse events. Additional details are reported in the summary of findings, Table 24.

Table 24 Children with ADHD: Summary of findings (GRADE)

Children with ADHD compared to placebo for iron deficiency without anaemia

Outcomes	№ of participants (RCTs)	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects*	
				Control group risk (only dichotomous outcomes)	Effect estimate (continuous outcomes) and risk difference (dichotomous outcomes) in children with ADHD
ADHD Severity	22 (1 RCT)	⨁◯◯◯ VERY LOW [a,b]	-	The mean ADHD Severity was **0**	MD **3.8 lower** (24.86 lower to 17.26 higher)
Clinical Global Impression Improvement	22 (1 RCT)	⨁◯◯◯ VERY LOW [a,c]	**RR 3.00** (0.19 to 47.96)	0 per 1'000	**0 fewer per 1'000** (0 fewer to 0 fewer)
RLS diagnosis	24 (1 RCT)	⨁⨁◯◯ LOW [a,d]	**RR 0.14** (0.04 to 0.45)	1,000 per 1'000	**860 fewer per 1'000** (960 fewer to 550 fewer)
Adverse events [e]	819 (7 RCTs)	⨁⨁◯◯ LOW [f,g,h]	**RR 1.12** (0.88 to 1.41)	385 per 1'000	**46 more per 1,000** (46 fewer to 158 more)
Serious adverse events [e]	828 (7 RCTs)	⨁◯◯◯ VERY LOW [i,j]	**RR 1.09** (0.35 to 3.37)	12 per 1'000	**1 more per 1'000** (8 fewer to 29 more)

*For dichotomous outcomes, the risk in the intervention group (and its 95% confidence interval) is based on the control group risk and the relative effect of the intervention (and its 95% CI). For continuous outcomes, the severity in the control group was not estimated.

CI: Confidence interval; MD: Mean difference; RR: Risk ratio

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

a. The RCT limitation was serious because selection bias (random sequence generation and allocation concealment) was unclear in 1 RCT; risk of attrition bias was high in 1 RCT and risk of selective reporting was unclear in 1 RCT.
b. Imprecision was very serious because the 95% confidence interval of the effect estimate was sufficiently wide to include appreciable harm or benefit when assuming an MCID of 15 points (half of a typical standard deviation from baseline value); this was consistent with the standardised effect estimate (-0.23, 95% CI -1.23 to 0.77) which also was sufficiently wide to include appreciable harm or benefit when assuming a medium effect of 0.5 SD. In addition, the total number of events was <300.
c. Imprecision was very serious because the 95% CI of the effect estimate is sufficiently wide to include both appreciable harm or benefit (relative risk increase or decrease greater than 25%) and because the total number of events was <300
d. Imprecision was serious because the total number of events was <300
e. Adverse events and serious adverse events were pooled over all populations (i.e. adults with RLS, women with fatigue and children with ADHD) and the quality of evidence was assessed across all study populations (seven RCTs).
f. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 4 RCTs and selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was unclear in 1 RCT; risk of detection bias was unclear in 4 RCTs; risk of attrition bias was unclear in 2 RCTs and high in 4 RCTs and risk of selective reporting was unclear in 3 RCTs and high in 1 RCT.
g. It was not downgraded for inconsistency because heterogeneity was low-moderate and confidence intervals were widely overlapping.
h. Imprecision was serious because the 95% CI of the effect estimate is sufficiently wide to include both no effect and appreciable benefit (relative risk increase greater than 25%) in favour of no iron therapy.
i. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 3 RCTs; risk of performance bias was unclear in 1 RCT; risk of detection bias was unclear in 5 RCTs; risk of attrition bias was unclear in 3 RCTs and high in 3 RCTs and risk of selective reporting was unclear in 4 RCTs and high in 1 RCT.
j. Imprecision was serious because the 95% CI of the effect estimate is sufficiently wide to include both appreciable harm or benefit (relative risk increase or decrease greater than 25%) in favour of no iron therapy and because the total number of events was <300.

2.4.4 Safety outcomes, all populations

Adverse events and serious adverse events were pooled across all three study populations due to the very low numbers (only seven RCTs reported safety outcomes). These seven RCTs reported a statistically non-significant increase in adverse events (critical outcome, low quality of evidence) and serious adverse events (critical outcome, very low quality of evidence) in patients treated with iron therapy compared to control.

Additional details are reported in the summary of findings, Table 25.

Table 25 Safety outcomes, all populations: Summary of findings (GRADE)

All populations compared to placebo for iron deficiency without anaemia

Outcomes	№ of participants (RCTs)	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects*	
				Control group risk (only dichotomous outcomes)	Risk difference (dichotomous outcomes) in all populations
Adverse events	819 (7 RCTs)	⨁⨁◯◯ LOW [a,b,c]	**RR 1.12** (0.88 to 1.41)	385 per 1'000	**46 more per 1'000** (46 fewer to 158 more)
Serious adverse events	828 (7 RCTs)	⨁◯◯◯ VERY LOW [d,e]	**RR 1.09** (0.35 to 3.37)	12 per 1'000	**1 more per 1'000** (8 fewer to 29 more)

*For dichotomous outcomes, the risk in the intervention group (and its 95% confidence interval) is based on the control group risk and the relative effect of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

a. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 4 RCTs and selection bias (allocation concealment) was unclear in 1 RCT; risk of performance bias was unclear in 1 RCT; risk of detection bias was unclear in 4 RCTs; risk of attrition bias was unclear in 2 RCTs and high in 4 RCTs and risk of selective reporting was unclear in 3 RCTs and high in 1 RCT.
b. It was not downgraded for inconsistency because heterogeneity was low-moderate and confidence intervals were widely overlapping.
c. Imprecision was serious because the 95% CI of the effect estimate is sufficiently wide to include both no effect and appreciable benefit (relative risk increase greater than 25%) in favour of no iron therapy.
d. The RCT limitation was serious because selection bias (random sequence generation) was unclear in 3 RCTs; risk of performance bias was unclear in 1 RCT; risk of detection bias was unclear in 5 RCTs; risk of attrition bias was unclear in 3 RCTs and high in 3 RCTs and risk of selective reporting was unclear in 4 RCTs and high in 1 study.
e. Imprecision was serious because the 95% CI of the effect estimate is sufficiently wide to include both appreciable harm or benefit (relative risk increase or decrease greater than 25%) in favour of no iron therapy and because the total number of events was <300.

2.5 Discussion

This section assessed the effectiveness of iron therapy from RCTs in three different iron-deficient, non-anaemic patient populations: adults with RLS, women with fatigue and children with attention-deficit hyperactivity disorder. The three populations will be discussed separately in the following three sections (2.5.1 - 2.5.3). Potential harm (safety outcome) was pooled across all three populations to increase the statistical power and will be discussed in the following section (see 2.5.4). Of note, health outcomes were not pooled across all populations (e.g. depression was reported for RLS and women with fatigue) because the specific beneficial effects of iron therapy in the different symptomatic populations was of interest.

The quality of evidence was assessed with GRADE for each specific outcome from the perspective of a systematic review author. Decision-makers and guideline authors, however, are advised to critically appraise the quality of evidence for all important and critical outcomes to make an overall rating of the quality of evidence as this is an iterative process. An overall rating may differ from the outcome specific ratings as presented in this report. Systematic review authors defined outcome specific thresholds (which are based on clinically important differences) to rate imprecision[21]. These thresholds should be carefully evaluated by decision-makers and may need to be adapted based on the balance and magnitude of the effects of other outcomes or based on decision-makers' values and preferences (for example, if both a clinically important benefit and a clinically important harm is shown for a PICO-question)[58,59].

2.5.1 Discussion – Adults with restless legs syndrome

RLS is a common neurologic syndrome for which the pathophysiology is not clearly understood. Possible explanations include alteration of dopaminergic function, brain iron metabolism or genetic factors[60,61], though causal pathways have yet to be determined[62]. RLS can be distinct into primary (unexplained) or secondary (associated with a comorbid condition). This assessment focused solely on the assessment of iron therapy on primary RLS. Iron deficiency is postulated as a cause of RLS patients, and thus iron therapy was suggested to be a potential treatment option for this condition[62]. Further potential pharmacological therapies of RLS include opioids, alpha-2-delta ligands and dopamine agonists[63].

Eight RCTs were included in the analysis of the clinical effectiveness of iron therapy in adults with RLS. There is weak evidence that iron therapy compared to control (mainly placebo/no iron and one trial pramipexole) decreases RLS symptom severity and improves quality of life and the global impression rating. The last two endpoints were not assessed in all trials. However, the overall quality of evidence according to GRADE was judged as very low, mainly because of methodological limitations of included RCTs (risk of bias), but also inconsistency, indirectness and imprecision. A potential "placebo effect" cannot be excluded in six out of seven trials reporting on RLS symptom severity.

Uncertainty of evidence is mainly due to attrition bias, and missing data with few indications of reasons. Most given reasons for incomplete follow-up and missing data were "lack of efficacy", "severity of symptoms", adverse events or "inconvenient treatment". Therefore, there is a high risk that effect estimates of health (RLS improvement) and safety outcomes (AE and SAE) might be biased due to missing data, underreporting and due to the relatively small number of events[27]. In addition, bias due to insufficient information for adequate allocation concealment in most of the RCTs is likely and result in an overestimation of treatment effects[64]. Because of potential selection bias and attrition bias, the quality of evidence had to be downgraded for most outcomes. Blinding of patients and study

personnel, however, was in most RCTs adequate and the risk of performance bias was considered to be low (quality of evidence not downgraded). Blinding of patients receiving oral iron most likely might have been insufficient due to stool colourisation of oral iron intake, but was not considered as a serious problem as formal blinding in trials of oral iron therapy was judged as correct.

For the majority of outcomes pooled, treatment effects from individual RCTs were homogenous. Inconsistency between individual studies (heterogeneity) was only observed for the outcomes sleep and global impression rating. Unexplained heterogeneity should decrease the confidence in effect estimates. The trial by Lee 2014 was the only included trial with an active control treatment with pramipexole, a dopamine agonist which may improve sleep quality[51]: In sensitivity analysis exclusion of Lee 2014 reduced heterogeneity remarkably for the outcome of sleep quality (see Figure 4). For global impression rating, heterogeneity couldn't be explored as only two studies reported this outcome. Inconsistency, therefore was judge to be serious and quality of evidence was downgraded.

In four RCTs with RLS patients, iron deficiency was not an inclusion criterion and in two RCTs rather a mixed population of patients with low to normal ferritin concentration was included. These RCTs were nevertheless included as no consensus for ferritin based iron deficiency thresholds exist and clinical judgement of iron deficiency-related symptoms may represent a sufficient indication for iron therapy. Two studies specifically excluded patients with very low serum ferritin concentration (<15 µg/L) and the results may not be generalizable to those with very low serum ferritin concentrations. Consequently, this deviation of the study population from the included RLS population compared to the population of interest was judged to be a serious problem leading to downgrading the quality of evidence due to indirectness. In addition, the heterogeneous trial population affects also the interpretation of the external validity of the found treatment effects (see section 2.6.1 of this report).

Imprecision (confidence in the effect estimate) was a general problem for almost all outcomes and further decreased the quality of evidence, mainly because the total sample size (OIS criterion not fulfilled) or the number of events was insufficient. Except for the outcome RLS symptom severity, the OIS was of borderline sufficiency. The 95% CI of the effect estimate (MD -4.23, 95% CI [-6.11, -2.34]) did not cross the line of no effect; therefore, the outcome was considered to be sufficiently precise, and the confidence in the effect estimate was adequate. On the IRLS severity scale, a clinically relevant difference was suggested to be around three[65,66], therefore, it could be interpreted that even a large majority of the study populations receiving iron therapy experienced an improvement of RLS symptoms compared to patients in the control group. In other words, the effect of iron therapy on RLS symptom severity is probably clinical relevant for a substantial proportion of RLS patients.

The pathophysiological pathways for postulated iron deficiency and RLS are unknown, but most likely involve dopaminergic neurotransmitter, brain iron metabolism and genetic factors[60,61]. Animal and autopsy studies have found markedly diminished iron and iron storage protein in the substantia nigra of RLS patients. This may support the theory that low brain iron plays probably a central role[60,62]. Findings from molecular PET and SPECT imaging study further illustrate in RLS patients the dysfunction of dopaminergic pathways that do not only involving nigrostriatal but also mesolimbic pathways[67]. In patients with iron deficiency the iron status is often quantified by measuring peripheral iron (e.g. serum ferritin), this, however, is a very poor proxy for brain iron status, because cerebrospinal fluid ferritin is poorly correlated with peripheral iron status[62]. In the present review, studies were identified with patients probably without peripheral iron deficiency, but it appears that these patients still benefit from iron therapy. Unfortunately, based on the present report, no conclusion about the

relation of the brain iron status of patients included in studies and iron therapy can be drawn, because this was not assessed in the respective studies. There was only one RCT (Earley 2009) which assessed also brain iron status[41]. Compared with the other included RCTs, Earley 2009 had a very invasive study design. In addition to the subjective measures (IRLS, etc.), the investigators collected blood and cerebrospinal fluid samples (lumbar puncture) and performed an MRI of the brain. The baseline assessment lasted for several days. After two weeks intervention, only serum ferritin concentration had significantly increased in the intervention group, and the increase in cerebrospinal fluid ferritin was only of borderline significance ($p = 0.04$). No significant change in the MRI iron index for the substantia nigra was reported. These findings, in combination with subjective outcomes, justified the premature stop of the study after two weeks because of "lack of both adequate power and any indication for clinically significant benefit"[41]. Hence, very little information could be gained from this RCT that was initially planned for two years follow-up period.

Four RCTs reported that participants stopped RLS concomitant treatment one to two weeks before the study began, in two studies concomitant RLS treatment was an exclusion criterion and in one study concomitant RLS treatment was not reported. Types of precedent RLS therapy were not reported. It is unclear to what extent a withdrawal of RLS treatment one to two weeks before study initiation affected the baseline measures. Only the study population from Davis 2000 was allowed to continue their RLS medications. It is unclear how concomitant RLS treatment affected iron therapy; Davis 2000 did not report on RLS severity and for the reported outcomes of sleep and QoL, iron therapy seemed to add no benefit (Figure 4 and Figure 6).

Lee 2014 was the only RCT included in the present report which compared iron therapy with a dopamine agonist (pramipexole). Based on this RCT, it appears that iron therapy has no effect on RLS severity when compared to pramipexole, whereas the other studies comparing iron therapy to placebo showed significant effects, except Earley 2009, which was prematurely stopped. It was decided to pool Lee 2014 with the other studies because sensitivity analysis showed that this study did not substantially alter the outcomes (see section 2.3.2.7).

Subgroup analysis on the route of administration did not reveal any differences between oral versus intravenous administration. Five of the eight RCTs that administered intravenous iron showed a statistical effect for RLS severity and were similar to the effect measured in one RCT (Wang 2009) using oral iron therapy. The second RCT administrating oral iron was Lee 2004 which showed no effect if iron when compared to pramipexole. The third RCT (Davis 2000) administrated iron orally, but did not report on RLS severity and was therefore not considered for subgroup analysis. For the other outcomes, especially AE, it is not possible to make any conclusions. As already described in earlier systematic overviews[68], it remains unclear which route of iron administration for patients with RLS is more effective and associated with fewer adverse events. Despite different administration frequencies of intravenous iron (single, double or multiple doses), the observed effects were consistent and CIs were widely overlapping, probably because the overall dose is the same for all five RCTs administrating intravenous iron. Two RCTs (Cho 2016, Trenkwalder 2017) administered 1000 mg iron in a single dose, two studies (Allen 2011, Earley 2009) administrated 500 mg iron at two different days and the fifth study (Grote 2009) administrated 200 mg iron at five occasions over a three-week period. Because the effects on RLS severity are similar, administration frequency does not seem to modify the effect estimates. However, Trenkwalder 2017, administrating iron on a single day, reported twice the number of adverse events in the intervention group, whereas Allen 2011 and Grote 2009, administrating iron

at two or at five days, reported similar numbers of adverse reactions/events in the intervention and control groups. The small number of trials precluded any conclusion of the observed treatment effects according to different iron regimes. Wang 2009, one of three studies administrating oral iron, also reported participants received vitamin C. Vitamin C is known to enhance gastrointestinal iron absorption[69] and might explain, at least partly, why the treatment effect observed by Wang 2009 are similar to the effects on RLS severity from the studies administrating iron intravenously.

Augmentation, the need to increase dopamine dosage over time to maintain the drug effect, is a typical problem in RLS patients taking this drug[66]. This problem has not been addressed in the one RCT that used a dopamine agonist pramipexole. The follow-up period of the included studies was rather short and does not allow any conclusion of long-term effects of iron therapy for RLS.

Comparison to existing literature

A Cochrane review by Trotti 2012[68] investigating iron therapy in RLS patients identified the same five trials (Allen 2011, Davis 2000, Earley 2009, Grote 2009, Wang 2009) and summary estimates for the reduction of restless legs syndrome severity were in the same direction as in the present report, but borderline non-significant (MD -3.79, 95% CI [-7.68, 0.10]). Findings for other outcomes like quality of life, sleep and adverse events were of similar magnitude although Trotti 2012 assessed the numbers of adverse events (instead of numbers of patients with adverse events). Withdrawal from the trial was used in the Cochrane review as a proxy for patient satisfaction with treatment which may be seen as problematic and not validated surrogate outcome.

Inclusion criteria in the present report allowed for other comparators than placebo, but only one study (Lee 2014) was identified that compared iron therapy with the dopamine agonist pramipexole. No meta-analysis comparing iron therapy with dopamine agonist was found; however, several, very recent systematic reviews[70-72] were identified which compared dopamine agonist to placebo or no treatment. Interestingly, the latest systematic review from Liu 2016 included 12 RCTs and reported very similar changes in improvement of IRLS severity (MD -4.64, 95% CI [-5.95, -3.33]) in dopamine agonists treated patients with as in the present report[71]. From the present report, no conclusion can be drawn regarding the efficacy of iron therapy vs. dopamine agonists.

2.5.2 Discussion – Women with fatigue

Fatigue is common in women of child-bearing age[73,74] and may manifest as a symptom in non-anaemic patients with iron deficiency[75]; although, the pathophysiological rationale appears not clear.

Four RCTs were included in the analysis of the clinical effectiveness of iron therapy in non-anaemic women with fatigue. Iron therapy compared to placebo improved fatigue severity and quality of life scores, but the overall quality of evidence was judged - according to GRADE – as very low, mainly due to study limitations (risk of bias), inconsistency and imprecision. A potential “placebo effect” cannot be excluded in the trials reporting on fatigue severity.

Attrition bias was unclear for two out of three RCTs reporting adverse and serious adverse events. Reporting bias would be most likely if the number of missing data outweighs the number of adverse events[27]. Hence, there is a high risk that safety outcomes are biased because of missing data or a likely reporting bias. In addition, for two RCTs no protocols were found, indicating an unclear risk of reporting bias. Because of these various biases, the quality of evidence was downgraded for the critical outcomes fatigue severity, adverse events and serious adverse.

Quality of evidence was further downgraded for heterogeneity, which was observed for the critical outcomes of fatigue improvement. Heterogeneity and inconsistency couldn't be explored because only two RCTs reported measures of fatigue improvement. Moderate heterogeneity (I^2=59%) was also observed among the three RCTs reporting adverse events. Quality of evidence due to inconsistency, however, was not downgraded because confidence intervals of single trials largely overlapped. Interestingly, in the FERRIM study more adverse events were observed in the control groups than in the intervention group whereas the two other studies reported more adverse events in the intervention group. The authors of the FERRIM study did not further elaborate on this issue.

Quality of evidence was downgraded for depression and anxiety because of indirectness due to the use of subscales of the "Current and Past Psychological Survey, CAPPS, which does not allow for an appropriate diagnosis of these outcomes. Depression and anxiety should be measured on a reliable and validated rating scale that is based on current the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) diagnostic criteria for depression and anxiety[76-78].

The confidence in effect estimate was insufficient for the outcomes of fatigue improvement, quality of life (total score), depression, anxiety, adverse events and serious adverse events because the sample size (OIS criterion not fulfilled) or the number of events was too small. However, the effect estimate was sufficiently precise for the critical outcome fatigue severity. The effect estimate for fatigue severity was reported on a standardised scale, whereas its 95% CI (SMD -0.41, 95% CI [-0.56, -0.26]) overlap the 0.5 SD threshold[16,23]. This 0.5 SD is a rule of thumb and a rather conservative approach to judge the clinical relevance. The 95% CI of the present effect estimate suggests that at least some patients receiving iron experienced a clinically relevant improvement compared to the control group. Although not directly comparable, the SMD is in line with the estimate deriving from the individual patient data meta-analysis (multilevel mixed linear regression model estimate: -0.36, 95% CI [-0.51, -0.12]).

The small number of trials also precluded a formal assessment for publication bias.

Patients in the four trials were quite similar in respect to age and extend of SF concentrations at baseline. However, the findings can only be generalised to women with serum ferritin concentrations below 50 µg/l because three RCTs recruited study subjects with a maximum serum ferritin concentration of 50 µg/l and one (PREFER) with a serum ferritin concentration below 15 µg/l. These differences in baseline serum ferritin concentrations did not materialize in relevant difference in effect sizes in subgroup analysis (see 2.3.3.6.2) probably because the median baseline ferritin concentration was around 15 µg/l in the PREFER trial and only slightly lower than in the remaining RCTs (17.3 to 22.0 µg/l, see Table 13 of the IPD meta-analysis), and hence, the difference at baseline was not substantial.

Likewise of the trial-specific subgroup analyses or the individual patient data meta-analysis using intravenous and oral iron administration did not reveal relevant differences in effect size for fatigue severity. However, the number of available RCTs and the study population was very limited to further explore how the route of administration affects treatment effects. It has been discussed that oral administration requires regular intake over several months which is associated with gastrointestinal side effects and patients' adherence[52]; moreover, adverse events due to oral iron are more frequent und of longer duration than adverse events due to intravenous iron[79]. In the present report, only one study (Vaucher 2012) administrating oral iron to women with fatigue reported adverse events and adverse events were common in the iron group, but the number of women experiencing gastrointestinal events was almost identical in both groups. The number of patients experiencing at

least one adverse event varied between studies and study arms (heterogeneity) and was not substantially lower with intravenous iron therapy. However, the number of events was too low for a substantive analysis. The follow-up period of four and 12 weeks of included trials and does not allow for any conclusions of intermediate or long-term effects of iron therapy in females with fatigue.

A recent Cochrane review from 2016 assessed RCTs of daily iron supplementation menstruating women with or without anaemia to improve anaemia, iron deficiency and health outcomes[80]. Fatigue was not a pre-specified outcome, but eight RCTs reported on fatigue, and five trials (including Verdon 2003) reported on fatigue improvement. The authors conclude that iron therapy appears to reduce fatigue symptoms.

A systematic review by Yokoi in 2017[75] assessed the effect of iron therapy on fatigue in patients with iron deficiency and no anaemia. Six RCTs were identified and a significant decrease in fatigue severity (pooled effect size 0.33, 95% CI 0.17, 0.48) - similar results to the present report (see Figure 18) – was found. However, Yokoi 2017[75] did not included the PREFER (Favrat 2014) trial but included three studies, two[81,82] of them being excluded in the present report due to non-randomised trial design and due to the inclusion of a non-symptomatic patient population[83]. Yokoi 2017[75], did not assess safety outcomes. In a systematic review by Houston 2018[84], the same trials as in the present report were selected for assessing the effectiveness of iron supplementation on fatigue severity and found an almost identical SMD of -0.38, (95% CI -0.52, 0.23). Houston 2018 did not pool other outcomes for the population of interest.

2.5.3 Discussion – Children with ADHD

Attention-deficit hyperactivity disorder (ADHD) usually manifests in childhood by impaired social functioning due to hyperactivity, impulsiveness and/or inattention[76-78]. Various genetic factors and neurotransmitter pathways have been identified to determine the pathophysiology of ADHD; nevertheless, the physiological processes and aetiology are not clear[85,86]. Additional factors such as diet or prenatal risk factors have also been put forward in the aetiology of ADHD, but evidence is limited[85-87]. Treatments range from behavioural interventions with/without combined pharmacotherapy to dietary interventions (elimination or supplementation).

Only one RCT of iron therapy in non-anaemic children with ADHD was identified. Iron-deficient children (18 boys and five girls) with serum ferritin <30 ng/ml) meeting the DSM-IV diagnostic criteria for ADHD ages five to eight years were included. The patient flow was not clear and ranged from a total of 17 to 19 children. The small number of events precludes any conclusion of benefit from iron therapy in this patient population. The overall quality of evidence according to GRADE was judged very low.

The study used the Conners' Parent Rating Scale (CPRS) and the Attention-Deficit Hyperactivity Disorder Rating Scale (ADHD RS) to measure ADHD at 12 weeks follow-up. Both scales are widely used tools for the measurement of ADHD and are validated for pre-school aged children measuring behaviours of ADHD and symptoms of ADHD according to the DSM-IV criteria[88,89].

No review on iron therapy in IDNA children with ADHD was found; however, two reviews on iron supplementation for ADHD (Cortese 2012[90] and Hariri 2015[91]) were identified in in iron-replete children. Both reviews were based on the same two trials (Sever 1997[92], Konofal 2008[57]) and were inconclusive due to low power and deficient trial methodology such as the inclusion of non-randomised trials (Sever 1997[92]).

2.5.4 Discussion – Safety outcomes, all populations

Study limitations were sufficiently described above and judged to be serious.

Across all patient populations, the 95% confidence interval for the pooled estimates was large and did not allow for precise estimates between treatment arms, and hence, imprecision was considered to be serious. Serious adverse events were rarely reported, probably because the populations were not seriously ill and iron therapy is not considered to cause serious side effects.

Adverse events were frequent, the number of patients that experienced an adverse event was 43% (180 out of 419) in the iron group and 39% (154 out of 400) in the placebo group. Interesting is the high number of patients with adverse events receiving placebo. A recent systematic review investigated the placebo and nocebo (number of adverse events in the placebo group) effect in the RLS population, the author reported that over 45% of the patients receiving placebo experienced at least one adverse event[93]. Importantly, a list of frequent adverse events was not presented in this report because the lists of adverse events in the individual RCTs were not reported (Vaucher 2012), were incompletely reported (PREFER, Trenkwalder 2017), or only selected adverse events were reported (FERRIM). Three RCTs (Allen 2011, Grote 2009 and Konofal 2008) reported a complete list of adverse events, however, those three RCTs reported together only 42 (12.6%) out of the total 334 adverse events and were therefore not considered to be representative.

The findings of the present report are in line with two Cochrane reviews which reported a similar non-significant effect of adverse events in a RLS patients (Trotti 2019: RR 1.48, 95% CI [0.97, 2.25]) and iron deficient menstruating women (Low 2016: RR 2.14, 95% CI [0.94, 4.86])[68,80,94]. A third systematic review of iron therapy in patients with fatigue did not report adverse events[75]. Identified systematic reviews children with ADHD did not mention or discuss adverse events[90,91].

Oral iron therapy is known to cause adverse gastrointestinal events[79]. In the present report, only two studies (Vaucher 2012, Konofal 2008) administered oral iron and reported adverse events. In the study by Vaucher 2012 the number of women with a gastrointestinal event was in both treatment groups the same (11.8% vs. 10.4%) and in Konofal 2008 too low to allow for meaningful comparisons. These observations are not in line with the Cochrane review by Low 2016 where a statistically significant increase in gastrointestinal adverse events with oral iron therapy was found with a suggestive iron dose dependency. A recent systematic review in anaemic patients with chronic kidney disease found no apparent difference in overall adverse events between oral or intravenous iron therapy, but revealed a statistically significant lower risk for adverse gastrointestinal from intravenous iron therapy (RR 0.43, 95% CI [0.28, 0.67])[95].

2.6 Conclusions

2.6.1 Conclusion - Adults with restless legs syndrome

The overall quality of evidence was judged to be very low because of the very low quality of evidence for the critical outcome serious adverse events.

The present report found a statistically and probably clinically relevant effect of iron therapy on the critical outcome RLS severity (low quality of evidence). Effects of iron therapy on the critical outcomes of sleep, adverse events, and serious adverse events were not statistically significantly different (all very low quality of evidence). The potential benefit from iron therapy for RLS severity reduction needs to be weighed against the slightly and statistically non-significantly increase in adverse events and serious adverse events from iron therapy and by considering the low and very low quality of evidence for the safety endpoints.

Generalizability of these findings for patients with RLS is very limited as the iron deficiency status of the trial populations was mainly unclear.

2.6.2 Conclusion – Women with fatigue

The overall quality of evidence was judged to be very low because of the very low quality of evidence for the critical outcome serious adverse events.

The present report found a statistically and probably clinically relevant effect of iron therapy on the critical outcome fatigue severity (moderate quality of evidence). Mental and physical health scores (important outcome, moderate qualities of evidence) were statistically significantly improved by iron therapy. Iron therapy compared to control showed no statistically significant improvement in overall quality of life scores (important outcome, moderate quality of evidence) and no statistically significant decreased depression scores (important outcome, low quality of evidence). Iron therapy compared to control showed a statistically significant improvement in anxiety scores (important outcome, low quality of evidence). The potential benefit from iron therapy for fatigue severity and other selected endpoints needs to be weighed against the slightly and statistically non-significantly increase in adverse events and serious adverse events from iron therapy and by considering the low and very low quality of evidence for the safety endpoints.

In addition, based on the available evidence from the individual patient data meta-analysis, no relevant associations were found between ferritin concentrations at baseline and the reduction of fatigue severity in women with fatigue.

The trial population of women with fatigue was homogenous and the findings of the present report can most likely be generalized to non-anaemic women with fatigue and ferritin concentration <50 µg/l, as it was defined by the inclusion criteria by the included RCTs. The use of different symptom scales for fatigue limits the interpretability and generalizability of reported summary estimates.

Based on the present report, no conclusion regarding the preferred route of administration, the impact of different serum ferritin concentrations at baseline on severity of fatigue, the effect in non-anaemic women with ferritin concentration above 50 µg/l and the other outcomes can be drawn.

2.6.3 Conclusion - Children with ADHD

The overall quality of evidence was judged to be very low because of the very low quality of evidence for the critical outcome serious adverse events.

No statistically significant or clinically relevant difference was found between iron therapy and placebo for the critical outcome of ADHD severity (very low quality of evidence). Additionally, other critical and important outcome measures were considered of low or very low quality of evidence, because the risk of bias for effect estimates was high and because the sample size of the one relevant study was very small. These limitations do not allow for any generalizations to a broader population of children with ADHD. Further studies should be undertaken to assess the effects of iron therapy on ADHD severity in non-anaemic, iron-deficient children with ADHD.

2.6.4 Conclusion – Safety outcomes

The overall quality of evidence was judged to be very low because of the very low quality of evidence for the critical outcome serious adverse events.

No statistically significant or clinically relevant difference was found between iron therapy and placebo for the critical outcome adverse events (low quality of evidence) and serious adverse events (very low quality of evidence). Adverse events were all considered to be mild in a large majority of cases and quickly resolved after intravenous injection, or after stopping oral treatment.

3 Cost-comparison and budget impact analysis

In the previous section, two symptomatic IDNA populations were identified which benefit from iron therapy when compared to placebo or control: Women with fatigue and adults with RLS. For these two populations, the SFOPH commissioned an economic evaluation for a comparison of parenteral versus oral iron therapy. The detailed rationale has been published in the scope[3].

3.1 Aim

The main objective of the cost-comparison and budget impact analysis was to quantify and compare the costs of parenteral and oral iron therapy from a health care payer perspective. The following two key research questions were addressed:

- What are the direct medical costs of oral iron therapy versus parenteral iron therapy in IDNA patients with fatigue or RLS?
- What is the budget impact of different iron treatment strategies in IDNA patients with fatigue or RLS?

3.2 Methods

3.2.1 Overview of the methodological approach

As a part of the scoping process, a systematic search was performed for economic studies and HTA reports on a direct comparison of oral versus parenteral iron therapy in IDNA patients[3]. This systematic search found no studies or reports focusing specifically on IDNA populations without any severe comorbidities, such as chronic heart failure, chronic or acute blood loss, or chronic kidney disease. Cost-effectiveness studies were either based on clinical RCTs of patients with such comorbidities, compared different brands of parenteral therapies without any reference to oral therapy, or did not report their results separately for IDNA patients (but rather for IDA or a mixed cohort of IDA and IDNA patients). Hence, none of these models could be directly adopted for the present assessment. Therefore, a new model was developed.

For the cost-comparison, the medical costs of all health care services of the different routes of iron administration were modelled with a decision tree reflecting the current clinical practice in Switzerland. The model was parametrized primarily with empirical evidence from the clinical trials identified in the section “Clinical effectiveness” of this HTA report (see section 2). However, as many variables required in the model were not reported in these trials, an extensive search of additional clinical literature was conducted (both RCTs and other study designs including empirical evidence of branch probabilities). In case of a lack of RCT-based, population-specific probabilities, data from other populations or settings were adopted. In case variables could still not be parametrized, clinical experts were asked for their best guess. Drug costs were based on prices from the “Spezialitätenliste (SL)”. Drug administration costs, as well as costs due to management of side effects, were based on TARMED positions, the “Analysenliste (AL)”, and the “Mittel- und Gegenständeliste (MiGeL)”. If inpatient treatment was a causal result of the iron therapy (e.g. due to a severe side effect), its costs were included based on SwissDRG case weights. Uncertainty was addressed by univariate, multivariate and probabilistic sensitivity analysis.

The budget impact analysis was based on the results from the cost-comparison analysis, epidemiologic data available for Switzerland and expert opinions.

3.2.2 Definition of the decision problem

3.2.2.1 Patients, intervention, comparator, outcome (PICO)

Empirical evidence from randomised controlled trials shows that iron therapy is effective compared to placebo in IDNA patients with symptomatic fatigue (in women) or with RLS (in adults of both genders). This is the result of the evaluation of the clinical effectiveness of this HTA report (see section 2). The economic evaluation therefore focused on these two populations. It compared the intervention of a first-line parenteral iron therapy with a first-line oral therapy. Within each of these first-line strategies, a switch from one to the other form of iron administration was possible during the course of the treatment. In accordance with the clinical experts advising this project, clinical practice shows that a share of the patients starting their therapy with a parenteral (/oral) treatment are switched to an oral (/parenteral) continuation before the completion of the therapy, due to side effects such as hypersensitive reactions, phlebitis, gastrointestinal problems or nausea.

According to the prescribing information of parenteral iron therapy (Ferinject®, Vifor Pharma and Venofer®, Vifor Pharma), parenteral iron therapy is indicated in patients where oral iron therapy was not effective or not tolerated, or in patients where oral iron therapy is contraindicated. However, according to the clinical experts advising this project, clinical practice in Switzerland shows that parenteral iron therapy is potentially chosen as first-line therapy, meaning that oral iron therapy was not tried first although it would have been indicated. Reasons for first-line parenteral therapy, mentioned by the experts, are diverse and may derive from the supply (physician) or the demand (patient) side. In order to reflect this current practice and in accordance with the SFOPH, both treatment strategies were considered to be relevant first-line therapies.

Two separate PICO structures, differing only in terms of the population, were initially defined for symptomatic fatigue and RLS, respectively. However, no evidence for different structures of the decision models, different branch probabilities, different resource uses, or different unit costs between the two populations was found. Consequently, the results presented for the cost comparison are applicable to both populations.

PICO 1: Women with fatigue

Population: IDNA women (at least 18 years of age) with fatigue and eligible for oral therapy

Intervention: Parenteral therapy with iron with possible switch to oral therapy

Comparator: Oral therapy with iron with possible switch to parenteral therapy

Outcome: Direct medical costs (drug costs, physician visits, drug administration costs, costs due to management of side effects both outpatient and inpatient)

PICO 2: Restless legs syndrome

Population: Adults (at least 18 years of age) with IDNA and with RLS and eligible for oral therapy

Intervention: Parenteral therapy with iron with possible switch to oral therapy

Comparator: Oral therapy with iron with possible switch to parenteral therapy

Outcome: Direct medical costs (drug costs, physician visits, drug administration costs, costs due to management of side effects both outpatient and inpatient)

3.2.2.2 Perspective

The cost comparison and the budget impact were modelled from a health care payer perspective. The present assessment took into account medical costs of all health care services (inpatient and outpatient) covered by the Swiss mandatory health insurance, irrespective of the actual payer (mandatory health insurance, other social insurance, government, out-of-pocket). The model did not include indirect costs due to reduced productivity and additional non-medical costs for patients, such as travel costs.

3.2.2.3 Structure of the decision model

A decision tree model including all treatment paths relevant to the economic evaluation over a time horizon of one year was designed. The tree initiates with the *decision node* of the model, as illustrated with a blue rectangle in Figure 31. This node marks the decision between the intervention (first-line parenteral iron therapy with a possible switch to oral therapy) and the comparator (first-line oral iron therapy with a possible switch to parenteral therapy). The physicians (or patients) make choices along the different paths of treatment, for example regarding the switch in the form of administration or regarding the termination/success of the therapy.

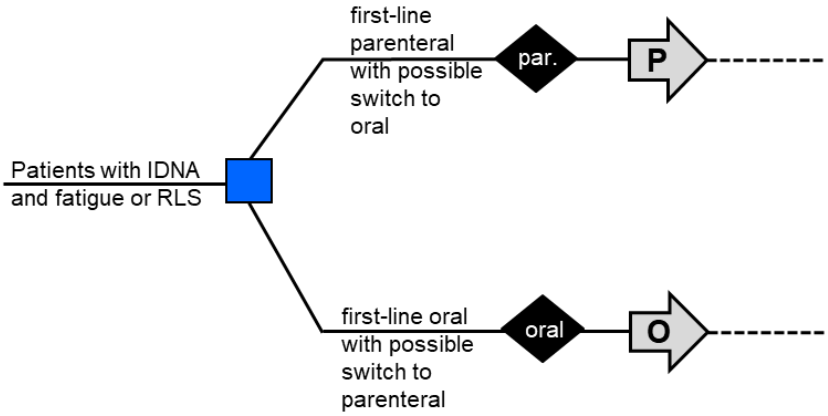
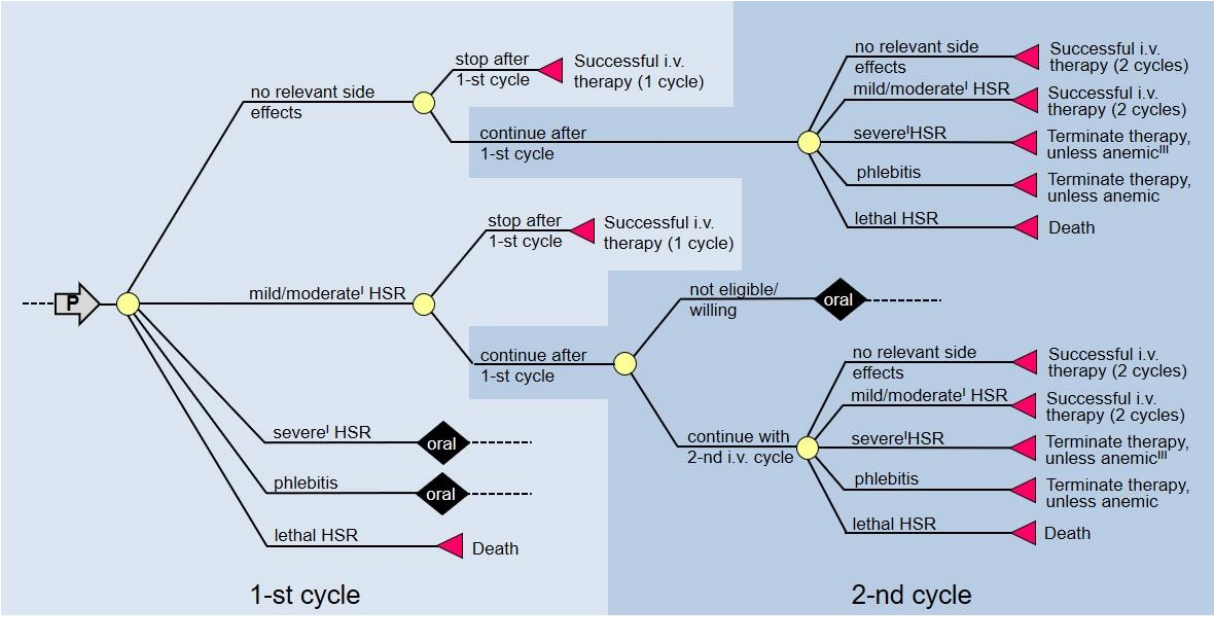


Figure 31 Decision between the relevant two routes of iron administration

Figure 32 illustrates the main branch of first-line *parenteral* therapy (the intervention therapy). This main branch leads to chance nodes (illustrated by yellow circles), sub-branches and endpoints (red triangles). The chance nodes indicate junctions in the decision tree at which patients follow different treatment pathways.



[I]Typology according to Rampton et al. (2014), "Hypersensitivity reactions to intravenous iron: guidance for risk minimization and management," Haematologica, 99(11).

Figure 32 Branch of first-line parenteral therapy

As a result of the iron infusion, each patient may experience one out of five possible adverse events:

- No relevant side effects caused by the parenteral therapy,
- a mild/moderate hypersensitive reaction (HSR) caused by the parenteral therapy, following the typology by Rampton et al. (2014)[96],
- a severe HSR,
- a phlebitis, and
- a lethal reaction.

Mild and moderate HSR include the symptoms of itching, flushing, urticaria, sensation of heat, chest tightness, hypertension, back/joint pains, and in the case of moderate HSR also cough, nausea, shortness of breath, and tachycardia[96]. Symptoms of a severe HSR are wheezing/stridor, periorbital edema, cyanosis, loss of consciousness, cardiac/respiratory arrest[96]. They can be life-threatening and exclude further parenteral iron therapy in patients with IDNA.

Following the top sub-branch of first-line parenteral therapy in Figure 32, with no relevant side effects of the first parenteral cycle, treatment is terminated either after the first cycle due to success (see definition below), or after a second cycle due to no success of the first circle (see below) or due to side effects during the second cycle, which lead to termination of iron therapy. Treatment success was defined as reaching a replete iron status (as judged by the treating physician) and/or having improved clinical symptoms. In the model, patients reaching a replete iron status with no clinical improvement after the first cycle were assumed not to suffer from any iron deficiency-related symptoms that would justify a second cycle and thus reached a terminal node. This is - in agreement with the consulted clinical experts - perfectly in line with clinical reasoning and management of these patient populations. A second cycle is indicated in cases with serum ferritin below the target level and no clinical improvement after the first cycle. Based on input from the clinical experts, parenteral iron therapy is

assumed to be successful with regard to blood parameters after the second cycle, regardless of whether the second cycle was accompanied by a mild/moderate HSR or not. Furthermore, it was assumed that a proportion of patients decides not to return to the physician at all after the first cycle (hence, they have lower costs due to no follow-up visit and laboratory tests).

If the parenteral cycle leads to a mild/moderate HSR, it was assumed that patients require additional supervision by the general practitioner (GP) and a prolonged infusion time (hence, they have increased costs compared to patients without relevant side effects), but receive the intended iron dosage (hence, they have the same success rate as patients without relevant side effects). This assumption slightly deviates from the recommendations by Rampton et al. (2014), who recommend that the administration should be terminated for some patients experiencing a mild/moderate HSR[96]. This assumption was made in agreement with the clinical experts, as three out of four treat their patients accordingly, whereas only one clinical expert stops the injection whenever such mild/moderate HSR occur. After experiencing a mild/moderate HSR in the first treatment cycle, iron therapy is either terminated due to success according to blood parameters, or the eligibility/willingness of the individual patient for another parenteral cycle is assessed. This means the physician evaluates eligibility of patients with mild/moderate HSR to undergo a second parenteral cycle and the patient has to evaluate his willingness to undergo a second parenteral cycle. If the physician or the patient decides against further parenteral therapy, the patient is switched to oral iron therapy. Otherwise, iron therapy is terminated after the second parenteral cycle with analogous reasoning as for the first sub-branch.

If at the beginning of the first parenteral cycle a severe HSR or phlebitis occurs, the administration is interrupted and the patient is switched to oral therapy.

Figure 33 illustrates the main branch of *first-line oral therapy* (the comparator therapy). The first junction takes into account that some patients interrupt oral therapy during the first cycle of treatment and are switched to parenteral treatment. According to the clinical experts advising this project, the reasons are mainly gastrointestinal side effects and nausea.

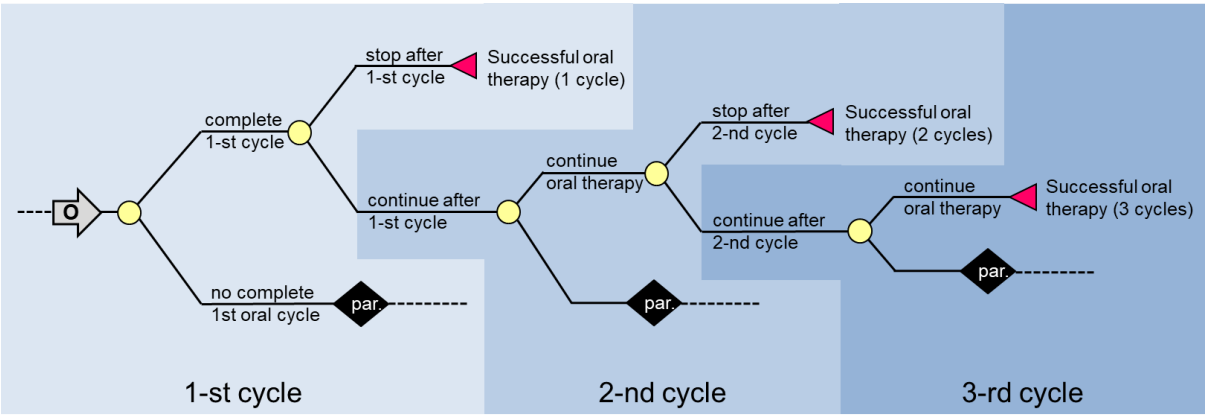


Figure 33 Branch of first-line oral therapy

If the first cycle of oral therapy is completed, a share of the remaining patients successfully terminate oral therapy, either by returning to the physician for a final follow-up blood analysis, which confirms treatment success, or by deciding not to return to the physician at all (hence, they have lower costs due to no follow-up visit and laboratory tests). The other share of the patients who complete the first

oral cycle either continue oral therapy with a second cycle or switch to parenteral iron therapy, due to unsuccessful oral treatment regarding the follow-up blood analysis. In the case of a second cycle with oral iron treatment, the therapy is either successful as proven by the follow-up blood analysis and therefore terminated (after a total of two cycles), or continued with parenteral therapy or a third oral cycle. As depicted in Figure 33, it is assumed that blood tests will show satisfying results latest after the third completed oral cycle. Further, it is assumed that after the second and third oral cycle, all patients have a follow-up blood analysis at the physician's, i.e. there are no *"no shows"*, as these are all patients who did return after the first cycle.

Summarizing the pathways of first-line oral therapy, each patient reaches one out of three endpoints, unless switched to parenteral therapy. The endpoints are: Successful therapy after one cycle of treatment (with or without follow-up visit and blood analysis), successful therapy after two cycles (with follow-up visit and blood analysis), and successful therapy after three cycles (with follow-up visit and blood analysis).

3.2.2.4 Definition of a treatment cycle

A treatment cycle comprises the initial physician visit, the evaluation of blood parameters, the medication, and the re-evaluation at the end of the cycle. The model was set up with a length per treatment cycle of three months, i.e. the patients' blood parameters are assessed three months after the initiation of the oral or parenteral treatment. According to the clinical experts, three months is a typical duration of an iron treatment cycle for oral iron therapy as well as for parenteral iron therapy in the patient population of interest.

This is supported by the following two publications from Switzerland:

- Fehr et al. (2009) recommend that patients receiving oral iron should be re-evaluated after a period of three months, and that after parenteral iron administration, re-evaluation should not be performed earlier than eight to 12 weeks[4].
- Martius (2009) recommends that re-evaluation of IDNA patients after parenteral therapy should not be undertaken before eight to 12 weeks after the administration, since before that, a strong but temporary increase in serum ferritin may occur[6].

As introduced in section 3.2.2.3, some patients require multiple treatment cycles, and some decide not to return to the physician for a follow-up visit after a successful therapy with symptom relief. By consequence, some of the sequential cycles do not include all of the above-mentioned components, such as the follow-up visits. In the case of parenteral iron therapy, some patients require a second administration a few weeks after the initiation of the cycle. This is modelled as an additional cost component of the respective cycle, while it does not affect its overall length of three months.

3.2.2.5 Time horizon

The time horizon of the analysis was one year. According to the clinical experts some patients require up to three treatment cycles (three three-month cycles, i.e. in total nine months) to fully recover. The underlying argument is that once iron treatment has been pursued by IDNA patients to this extent, they are expected to show satisfying blood parameters, implying that the treatment was successful. Otherwise, if a patient turns anaemic or suffers from comorbidities, which prevent the iron therapy from achieving its purpose, the patient does not comply with the defined PICO. Further, if the blood

parameters are indeed satisfying, but the patient still claims symptoms of fatigue or RLS, other conditions than iron deficiency must be assumed to cause the symptoms.

For RLS patients it was not possible to make assumptions based on expert experience but it was considered that a similar time horizon can be applied given that the RCTs in this indication used similar treatment durations as the RCTs investigating fatigue (oral iron treatment for 24 weeks was the maximum in the RCTs investigating RLS). Consequently, for both populations a time horizon of one year was considered long enough to model all relevant consequences related to the initial decision regarding first-line treatment strategy.

Note that switching from oral to parenteral therapy and vice versa, as illustrated in Figure 32 and Figure 33, was not modelled by moving through the respective branch of oral/parenteral therapy from beginning to end. A patient completes a maximum of three cycles, irrespective of whether the patient switches treatment or not. If, for example, a patient switches to parenteral treatment after two cycles of first-line oral treatment, only one cycle of parenteral treatment is modelled, in order to end with a total number of three cycles (two oral, one parenteral). Furthermore, a patient cannot return to the initial first-line form of iron administration, once it has been abandoned for the other form. This doesn't imply that such pathways are impossible in practice, but it was assumed that they occur with small probabilities and therefore have no relevant influence on the cost comparison.

3.2.2.6 Discounting

The cost comparison analysis covered a time horizon of one year. It was therefore refrained from discounting.

3.2.2.7 *Cost types*

Direct medical costs of oral and parenteral iron therapy were compared from a health care payer perspective. The present assessment took into account medical costs of all health care services (inpatient and outpatient) covered by the Swiss mandatory health insurance, irrespective of the actual payer (mandatory health insurance, other social insurance, government, out-of-pocket). Inpatient treatment may occur as a result of side effects during parenteral iron therapy. It was assumed that side effects due to oral iron therapy do not lead to inpatient treatment. Indirect costs, such as productivity losses, as well as direct non-medical costs such as travel expenses were not accounted for in the analysis because of the perspective chosen. All costs are reported in Swiss Francs for the year 2018.

3.2.3 Data sources for the parametrization of the model

This assessment aimed at evaluating the daily routine of general practitioners in Switzerland. However, there exists very limited scientific evidence regarding the Swiss practice and no binding treatment guidelines are in place. Consequently, several model input parameters had to be based on expert opinion. Expert opinion was gained from the clinical experts recommended by the SFOPH.

While fatigue and RLS are different populations, no empirical evidence could be found of these two conditions differing in terms of their pathways of iron treatment. Therefore, and in accordance with the suggestions by the clinical experts, it was assumed that the branch probabilities and cost components do not differ across the two symptomatic groups. The cost comparison between oral and parenteral iron therapy was thus performed using the same model.

3.2.3.1 Branch probabilities

The data populating the branch probabilities of the decision tree were extracted from published reports of clinical studies, such as RCTs, non-randomized clinical studies, and retrospective studies, with the type depending on the probability of interest. Whenever the required information was reported in the studies identified in the section "Clinical effectiveness" of this HTA report (see section 2), branch probabilities were extracted from this source. However, as many probabilities required in the model were not reported in these RCTs, an extensive search of additional clinical literature was conducted. The detailed search strategy, in- and exclusion criteria for branch probabilities can be found in the appendix (details see Appendix 5.4 and 5.6). Details about the branch probabilities for side effects are described in Appendix 5.8.

In case of a lack of RCT-based, population-specific (IDNA) probabilities, data from another population or setting were adopted. This was the case for the branch probabilities of experiencing a severe HSR, a lethal HSR, or a phlebitis. They were extracted from studies which either do not distinguish between anaemic and non-anaemic patients or which explicitly concern anaemic patients. Hence, the assumption was made that these probabilities are the same for IDNA and IDA patients (details see Appendix 5.4).

In case probabilities could still not be found in the published literature, the clinical experts were asked for their best guess based on their practical experience. These inputs were not only important to obtain branch probabilities, but also to validate the pathways in the decision tree.

3.2.3.2 Resource use

Based on input from the clinical experts, several assumptions in regard to the application of drugs were made which are detailed in Appendix 5.7.1.

3.2.3.3 Costs per unit

Drug costs for oral and parenteral iron treatment as well as for the treatment of phlebitis were estimated based on official drug prices available from the 1st of July 2018 from the specialities list issued by the SFOPH[97]. For the oral medication, an equal market share of all the six products available (Duofer®, Ferro sanol®, Ferrum Hausmann®, Maltofer®, Tardyferon®, Kendural®) was assumed and used the average of the prices for 1 mg of substance for the largest packets per product. The same unit price was calculated for the two drugs available in the parenteral treatment, and they were weighted by their market share according to a report by Helsana insurance company (Ferinject® 86.3%, Venofer® 13.7%)[98]. Details can be found in the appendix (details see Appendix 5.7.2 and 5.7.3).

For the office visits fee-for-service rates according to the Swiss medical tarif code for outpatient services (Tarmed) were applied[99]. It was assumed that the treatment, as well as the follow-up consultations, take place at the general practitioner (positions: 0.0010, 0.0020, 0.0030, 0.0855, 0.137; details see Appendix 5.7.4). The costs of the consultations were calculated by multiplying the resulting tax points according to Tarmed with a weighted average (the weight was given according to the number of general practitioners in each canton based on data from the FMH) of the tax point values set by the cantons[100]. Tarmed was also used to estimate costs of blood sampling for lab tests (position: 0.0715).

Material costs related to parenteral medication were based on the "Mittel- und Gegenstände-Liste (MiGeL)" issued by the SFOPH (positions: 03.04.01.00.1, 03.04.04.00.1, 03.04.05.00.1 and 99.11.01.00.1; details see Appendix 5.7.5)[101].

Unit costs for the laboratory tests were taken from the Analysenliste issued by the SFOPH (positions: 1370.00 (hemogram), 1314.00 (ferritin))[101]. It was assumed that the hemogram is performed by the GP and ferritin is measured in a private laboratory. Therefore, position 4700.00 was used in addition for estimating the costs of measuring ferritin (details see Appendix 5.7.4).

Severe HSR was assumed to be treated in a hospital inpatient setting. For the ambulance transport from the GP to the hospital, the costs from the study by Wieser et al. (2012) were used[102] and were adjusted for inflation as measured in the subcategory "outpatient services" available from "Landesindex für Konsumentenpreise[103]" to estimate costs for 2018 (CHF 1'618 * (100.0684/95.8961) = CHF 1'688). For estimating the costs for the inpatient treatment, the average costs per case for the Swiss DRG (diagnosis-related group) X60B in 2014 available from the Federal Statistical Office[103] were used and adjusted for inflation as measured in the subcategory "inpatient hospital services" available from "Landesindex für Konsumentenpreise[103]" to estimate costs for 2018 (CHF 4'227 * (96.6241/100.8145) = CHF 4'051). The DRG X60B was obtained by grouping a case with ICD-10 T88.6 in the SwissDRG grouper[104]. Total costs for an inpatient treatment of a severe HSR including ambulance transport was therefore estimated at CHF 5'740.

The costs of lethal HSR were assumed to be the same as for the inpatient treatment of a severe HSR (CHF 5'740).

3.2.3.4 Costs per component

Based on the resource use and costs per unit described above, costs for the different components of the model were calculated. These costs are summarized in Table 26. The last column refers to costs of the component as described in the first column.

Table 26 Overview on cost components

Component	Source	Comments	Component costs (CHF)
Drugs: Oral therapy, 1 cycle (3 months)	Specialities list[97]	Average of the lowest price for 1 mg per available drug (biggest package size); base case: 100mg/day; 90 days For details of the calculation, see Appendix 5.7.2.	30.57
Drugs: Parenteral therapy, 1 infusion	Specialities list [97], Helsana Arzneimittelreport[105], clinical experts	Average of the lowest price for 1 mg per available drug (biggest package size), weighted by market share[105]; base case: 500 mg per visit For details of the calculation, see Appendix 5.7.3.	160.68
Parenteral therapy: GP visit required for infusion and material per infusion	Tarmed[99], MiGeL[101], clinical experts	Consultation (base case: 10 min) by the GP and surveillance of the infusion (base case: 30 min)	124.76
Follow-up: GP visit	Tarmed[99]	Base case: 15 min	41.15
Follow-up: Laboratory tests	Analysenliste[101]	Hemogram (Pos. 1370.00) and Ferritin (Pos. 1314.00); base case: 20% of the patients	40.76

		hemogram only, 80% hemogram and ferritin	
Side effect treatment: mild/moderate HSR per case	Tarmed[99]	Additional time needed: 5 min consultation + 15 min surveillance	44.91
Side effect treatment: severe HSR per case (inpatient and ambulance transport)	Statistik diagnosebezogener Fallkosten 2014[103], Wieser et al., 2012[102]	DRG X60B	5'740
Side effect treatment: Phlebitis per case	Tarmed[99], specialities list[97]	1 extra visit at the GP, 1 package of Ibuprofen, 1 package of Venugel	62.25
Side effect treatment: lethal HSR per case (inpatient and ambulance transport)	Statistik diagnosebezogener Fallkosten 2014[103], Wieser et al., 2012[102]	DRG X60B	5'740

3.2.4 Sensitivity analysis

A number of univariate sensitivity analyses were performed. The impact of variations in some of the input parameters with a high degree of uncertainty were assessed, i.e. all the branch probabilities, resource use and some of the unit costs. In the univariate case, each parameter was varied one by one, setting it to its lower and upper bound, respectively, while leaving all the other parameters at their base case value. This procedure allows for the identification of the most important single impact factors on the cost estimates. The upper and lower bounds used in the univariate sensitivity analysis can be found in Table 27.

A number of two-way sensitivity analyses were also performed. In this analysis, two factors that showed a high impact in the univariate sensitivity analysis were simultaneously varied and the impact on changing both variables on the cost difference between parenteral and oral iron therapy was assessed.

To further assess uncertainty, a probabilistic sensitivity analysis was also performed. In this analysis, all input parameters analysed in the univariate sensitivity analysis were varied randomly at the same time [106]. As the lower and upper bounds of probabilities used in the univariate sensitivity analysis stem from different studies or represent different expert opinions, uniform distributions were deemed as appropriate. As uncertainty behind dosages, duration of GP visits, costs of laboratory tests and costs of treating severe HSR has the same source as mentioned for the probabilities above, these parameters were also simulated to follow uniform distributions. The model was run 10'000 times.

3.2.5 Budget impact analysis

3.2.5.1 Estimating the target population

For the budget impact analysis, the eligible patient population was estimated first[107]. This population corresponds to the number of adult patients with IDNA (fatigue/RLS) and treated with iron within one year. Several studies from Switzerland that report the prevalence of iron deficiency were identified[108-111]. However, none of these studies reported whether the patients had any symptoms or whether they were treated. The study by Biétry et al. (2017) retrieved data from Helsana (one of the largest Swiss health insurance companies) to estimate the use of iron therapy[98]. They included all patients with at

least one prescription for a drug coded in the anatomic therapeutic chemical classification system (ATC) as class B03A (oral and parenteral iron drugs; multivitamins were excluded). Furthermore, they excluded all patients with a diagnosis of cancer. They reported a 3-year prevalence of 16.0% for women and 2.6% for men for 2012-2014. This 3-year prevalence was divided by three to approximate a 1-year prevalence and accounted for the 0.3%-point increase from 2012 to 2014. Assuming a linear progression, this increase is equal to an annual increase of 3.3% (0.15%/4.6%). Therefore, the 1-year prevalence in women was estimated at 5.16% for 2012, 5.33% for 2013 and 5.51% for 2014. For men the prevalence was estimated at 0.84% for 2012, 0.87% for 2013 and 0.89% for 2014. It was assumed that this trend continued up to 2018 and consequently, the 1-year prevalence in 2018 was estimated at 6.3% for women and 1.0% for men. This number reflects iron therapy in general. Therefore, additional information was needed to estimate the prevalence of symptomatic (women with fatigue and women or men with RLS) IDNA.

The clinical experts were asked for their estimation of the share of patients treated with iron in one year that are treated due to iron deficiency anaemia. Furthermore, they were asked for an estimation of the percentage of patients treated with iron due to IDNA with symptoms other than fatigue/RLS. These estimations allowed for calculating the target population for the budget impact analysis. Two experts felt not comfortable enough to give any estimations and the estimations given by the other two experts varied widely. Therefore, a scenario based on the mean of the two expert opinions was also calculated.

For data on the population size, the latest statistics from the Federal Statistical Office for the end of the year 2017 were used (which is equivalent to the beginning of 2018)[112].

The uncertainty for the target population for the reference year 2018 was high, therefore, it was deemed not appropriate to make any projections regarding the future target population. Consequently, future population changes and potential changes in the disease awareness were not taken into consideration.

3.2.5.2 Treatment mix

Both treatment strategies are currently used in Switzerland. In the study by Biétry et al. (2017), oral iron therapy had a prevalence of 3.4% and parenteral iron therapy 1.9% in 2014[98]. With the 1-year iron deficiency prevalence of 4.5% also reported by Biétry et al. (2017), this would mean that 0.8% (3.4% + 1.9% - 4.5%) are treated with both oral and parenteral iron, 2.6% (3.4% - 0.8%) are treated with oral iron only and 1.1% (1.9% - 0.8%) with parenteral iron only. This means that 24.4% (1.1%/4.5%) of iron deficient patients are treated with parenteral iron only. For the budget impact estimation a situation, where 0% of the patients receive first-line parenteral, was compared to a situation where 24.4% of the patients receive first-line parenteral. A hypothetical maximum of the budget impact was also estimated by comparing a situation where 0% of the patients receive first-line parenteral, to a situation where 100% of the patients receive first-line parenteral.

For the costs per patient treated with first-line parenteral and first-line oral, respectively the base case result from the cost-comparison analysis were used primarily. As this result is subject to a substantial amount of uncertainty, the estimations from the probabilistic sensitivity analysis were also taken into consideration. The 95% lower bound of the cost difference was used to estimate a lower bound of the budget impact (scenario “minimum cost difference between treatment strategies”) and the 95% upper

bound of the cost difference was used to estimate an upper bound of the budget impact (scenario “maximum cost difference between treatment strategies”).

3.2.6 Technical implementation

The model including sensitivity analyses was implemented using Microsoft Excel 2016.

3.3 Results

3.3.1 Branch probabilities

Branch probabilities were extracted, whenever possible, from the studies identified in the section "Clinical effectiveness" of the present report (see section 2). However, as many probabilities required in the model were not reported in these RCTs, an extensive search of additional clinical literature was conducted. The detailed search strategy, in- and exclusion criteria for branch probabilities can be found in the appendix (details see Appendix 5.4 and 5.6). The parametrization of the base case model, as well as the lower and upper bound used for the sensitivity analysis, are listed in Table 27. Therein, the first column indicates the respective chance node, which were numbered to increase orientation within the model (see Figure 34 and Figure 35).

Table 27 Branch probabilities of the decision tree

# node	Description	Source	Base case value	Sensitivity analysis	
				Lower bound	Upper bound
Parenteral therapy					
1a	Probability of experiencing a mild/moderate HSR	Favrat B, et al. (2014)[53]; Krayenbuehl PA, et al. (2011)[52]; Trenkwalder C, et al. (2017)[46]	27.6%	20.9%	31.0%
1b	Probability of experiencing a severe HSR	swissmedicinfo.ch (Ferinject®)[113];	0.5%	0.1%	1.0%
1c	Probability of experiencing phlebitis	Broche DE, et al. (2005)[114]; Quintana-Diaz M, et al. (2017)[115]; Diez-Lobo AI, et al. (2007)[116]	2.3%	0.4%	6.5%
1d	Probability of experiencing a lethal HSR	Rampton D, et al. (2014)[96]; Chertow GM, et al. (2006)[117]	0.00002%	0.000012%	0.000078%
2	Probability of stopping therapy after first parenteral cycle	Clinical experts	90.0%	85.0%	95.0%
3	Probability of not being elligible for second parenteral cycle	Clinical experts	5.0%	2.5%	7.5%
Oral therapy					
4	Probability of completing first oral cycle	Suominen P, et al. (1998)[118]; Zaim M, et al. (2012)[119]; Paesano R, et al. (2010)[120]	87.8%	84.5%	91.0%
5	Probability of stopping therapy after first oral cycle	Clinical experts	85.0%	80.0%	90.0%
6	Probability of continuing oral therapy after first oral cycle	Clinical experts	90.0%	85.0%	95.0%
7	Probability of stopping therapy after second oral cycle	Clinical experts	95.0%	92.5%	97.5%
8	Probability of continuing oral therapy after second oral cycle	Clinical experts	99.0%	95.0%	100.0%

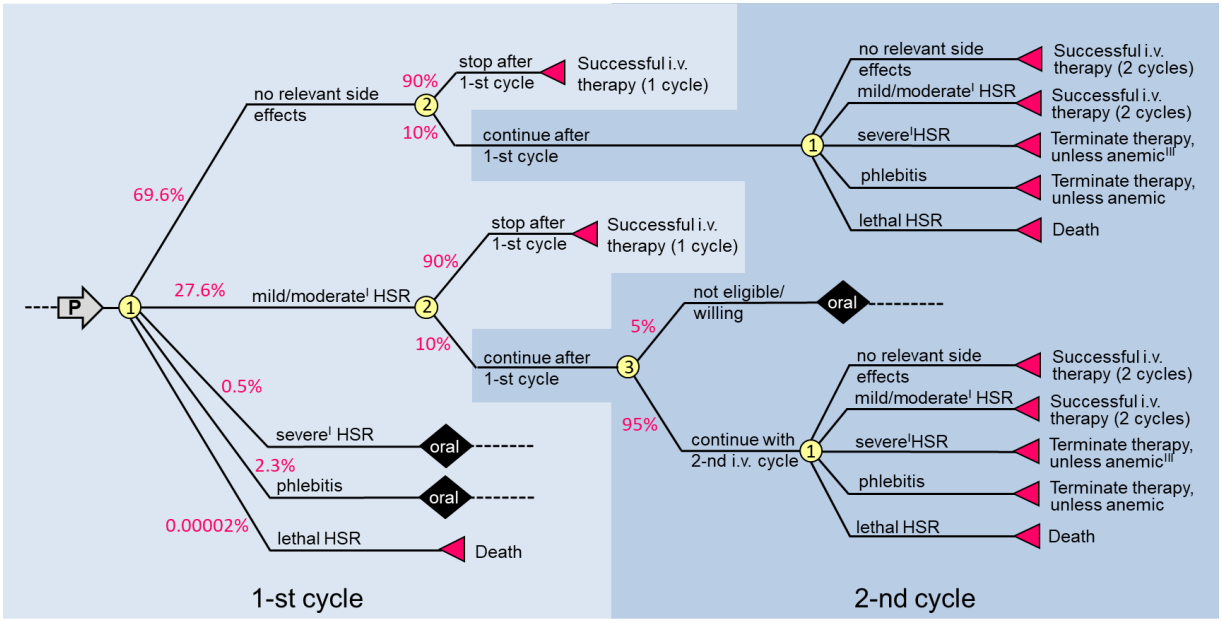


Figure 34 Branch of first-line parenteral therapy including probabilities for base case analysis

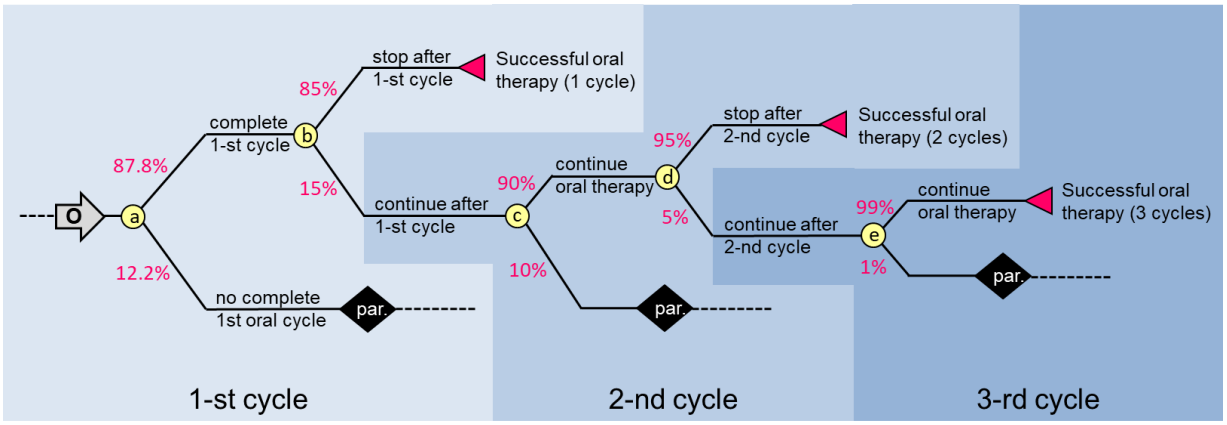


Figure 35: Branch of first-line oral therapy including probabilities for base case analysis

Different categorizations of adverse events and different levels of details were reported in the different studies. In addition, the adverse events were compared to the ones reported by Rampton et al. (2014)[96] to judge which events potentially qualify for a mild/moderate HSR according to the assessment. The details are provided in Appendix 5.8.

3.3.2 Validation of the model

The base case calibration of the model implies that **87.5%** of the patients with a *first-line parenteral* treatment strategy experience treatment success within the first cycle (Table 28). Thereof, 62.8 percentage points finish the first parenteral cycle with no relevant side effects, and 24.8 percentage points experience a mild/moderate HSR but nevertheless achieve treatment success. If a *first-line oral* treatment strategy is pursued, **85.3%** of the patients experience treatment success within the first cycle of treatment. 74.6 percentage points of these patients have a successful oral therapy without side effects. The other fraction of 10.6 percentage points are patients with side effects due to oral

therapy, who therefore switch during the first treatment cycle and experience treatment success as a result of parenteral therapy. These results are consistent with the fractions suggested by the clinical experts in the scope of the economic analysis, where it was indicated that between 80% and 90% of the patients are successfully treated within the first cycle.

Table 28 Probabilities of pathways leading to success after one cycle of treatment (three months)

Pathways	**Probability (cumulative)**	**Probabilities along the pathway** (see Table 27 and Figure 34 and Figure 35)
First-line parenteral		
Experiencing no relevant side effects and stop after first parenteral cycle	62.6%	0.696 x 0.9
Experiencing mild/moderate HSR and stop after first parenteral cycle	24.8%	0.276 x 0.9
Total	87.5%	
First-line oral		
Complete first oral cycle and stop after this first cycle	74.6%	0.878 x 0.85
No complete first oral cycle, switching to parenteral, experiencing no relevant side effects and stop after first parenteral cycle	7.6%	0.122 x 0.696 x 0.9
No complete first oral cycle, switching to parenteral, experiencing mild/moderate HSR and stop after first parenteral cycle	3.0%	0.122 x 0.2076 x 0.9
Total	85.3%	

Treatment success is achieved after *two cycles* of treatment among **11.5%** of the patients with a *first-line parenteral* treatment strategy. 9.3 percentage points out of these 11.5% refer to patients with two sequential parenteral treatment cycles, with no side effects or with a mild/moderate HSR (during the first, the second, or both cycles). The remaining 2.2 percentage points encompass patients with a successful oral treatment cycle, after a parenteral treatment cycle leading to a mild/moderate HSR, a severe HSR, or phlebitis. Of the patients with a *first-line oral* treatment strategy, **13.5%** achieve treatment success after two cycles. 11.3 percentage points thereof represent patients with two sequential oral treatment cycles. The other 2.3 percentage points of patients have either a mixed pathway, with an oral cycle followed by a parenteral cycle, or two parenteral cycles after the first oral cycle had been interrupted.

In summary, among both first-line parenteral treatment strategy, and first-line oral treatment strategy, the probability of patients to achieve treatment success *within the first two treatment cycles (three or six months*) amounts to **99%** (More precisely, the proportion amounts to 99.0% for the first-line parenteral treatment strategy, and to 98.9% for the first-line oral treatment strategy.). In the case of *first-line parenteral* therapy, 96.8% of patients achieve treatment success only being treated with the parenteral route of administration until the end of the three (/six) months, with 29.3%-points experiencing a mild/moderate HSR at least once. In the case of *first-line oral* therapy, 85.9% of patients attain treatment success only being treated with the oral route of administration for three (/six) months. In both first-line therapy schemes, only 1% of patients require three treatment cycles.

3.3.3 Base case results

The following costs of the first-line parenteral therapy (intervention) and of the first-line oral therapy (comparator) are based on the calculations of the decision tree with the probabilities and costs as derived above. They refer to total (direct) medical costs for the time horizon of one year considered in each treatment strategy.

The estimated medical costs for the first-line parenteral therapy are **CHF 561** per patient. For the first-line oral therapy, they amount to **CHF 182**. The difference in costs between the two treatment strategies is therefore estimated to be **CHF 379** per patient.

3.3.4 Univariate sensitivity analysis

A univariate sensitivity analysis was performed to modify the input data within a plausible range. Table 29 displays the lower and upper bounds used in the univariate analysis and the results obtained by the simulation. The numbers in the last column refer to the differences to the base case result, not the difference between the two treatment strategy's costs in the respective scenario. All the probabilities from the decision tree and relevant utilization and cost parameters were varied.

The univariate sensitivity analysis shows the effect of changing one parameter at once to its lower and upper bound, respectively, while leaving all the others at their base case value. The Tornado diagram (Figure 36) shows the effect of each univariate change on the difference in total costs between both treatment strategies.

The dosage of the parenteral administration per visit clearly has the biggest impact on the difference in total costs per patient (+/-21.2% compared to the base case difference). It is followed by the visit duration for a parenteral treatment (+14.8%; no lower bound defined). The third largest effect has the probability of experiencing a severe HSR (-5.4%; +6.4% compared to the base case difference).

The smallest effect on the cost difference is caused by the probability of experiencing a lethal HSR after a parenteral treatment (+/-0% compared to the base case difference).

Table 29 Parameter inputs for univariate sensitivity analysis

Parameter Description		Base case	Lower and upper bound in univariate sensitivity analysis	Result (difference in CHF to base case)
1a	Probability of experiencing mild/moderate HSR	27.6%	[20.9%; 31.0%]	[-0.11; 0.06]
1b	Probability of experiencing severe HSR	0.5%	[0.1%; 1.0%]	[-21.69; 27.09]
1c	Probability of experiencing phlebitis	2.3%	[0.4%; 6.5%]	[-0.25; 0.54]
1d	Probability of experiencing lethal HSR	0.00002%	[0.000012%; 0.000078%]	[-0.00; 0.00]
2	Probability of stopping therapy after first parenteral cycle	90.0%	[85.0%; 95.0%]	[22.58; -22.58]
3	Probability of not being elligible for second parenteral cycle	5.0%	[2.5%; 7.5%]	[0.24; -0.24]
4	Probability of completing first oral cycle	87.8%	[84.5%; 91%]	[-15.98; 15.49]
5	Probability of stopping therapy after first oral cycle	85.0%	[80.0%; 90.0%]	[-7.74; 7.74]
6	Probability of continuing with oral therapy after first oral cycle	90.0%	[85.0%; 95.0%]	[-3.00; 3.00]
7	Probability of stopping therapy after second oral cycle	95.0%	[92.5%; 97.5%]	[-0.34; 0.34]
8	Probability of continuing oral therapy after second oral cycle	99.0%	[95.0%; 100.0%]	[-0.10; 0.02]
9	Probability of returning for follow-up visit after oral cycle	80.0%	[70.0%; 90.0%]	[5.93; -5.93]
10	Daily dosage of oral medication	100mg	[80mg; 150mg]	[6.67; -16.67]

11	Dosage of one i.v. infusion	500mg	[300mg; 700mg]	[-85.27; 85.27]
12	Duration of follow-up visit	15 min	[10 min; 20 min]	[1.37; -1.37]
13	Duration of i.v. infusion visit (consultation + surveillance)	10 min + 30 min	[lower bound not applicable; 30min±15 min*]	[-; 59.59]
14	Labtests performed (hemogram only/ferritin only/ combination)	20%/0%/80 %	[100%/0%/0%; 0%/0%/100%]	[2.12; -0.53]
15	Cost of treating severe HSR	CHF 4'205	[-30%;+30%] = [2'943; 5'466]	[-8.16; 8.16]

**This Tarmed-position (0.137) has a unit of 15 min and was therefore varied to 30 min or 45 min.*

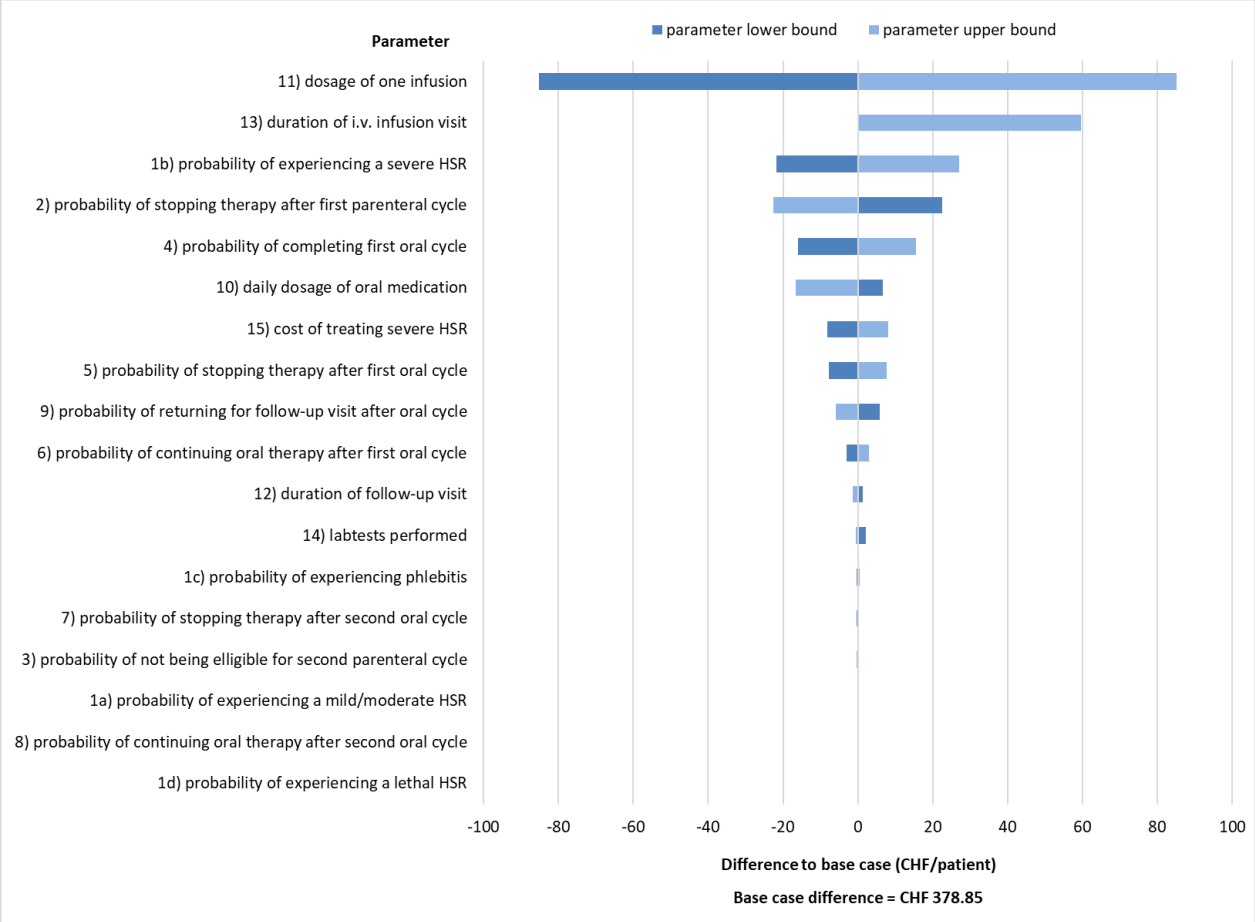


Figure 36 Tornado diagram showing the impact of a univariate change of single parameters on the result

3.3.5 Multivariate sensitivity analysis

Three different two-way sensitivity analyses were performed in which two parameters were allowed to vary at the same time (details in section 3.2.4). The three combinations were chosen according to the magnitude of their influence in the univariate sensitivity analysis:

- Dosage of parenteral medication administered in one session and probability of a severe HSR
- Dosage of parenteral medication administered in one session and probability of stopping therapy after first parenteral cycle
- Probability of a severe HSR and probability of stopping therapy after first parenteral cycle

The parameter “duration of i.v. infusion visit”, which showed the second biggest impact in the univariate sensitivity analysis, was not included in the multivariate sensitivity analysis as this parameter could only be varied in one direction.

The parameters were varied within the range defined by their lower and upper bound, resulting in a 3x3 matrix of results for each combination of parameters. Results are depicted in Figure 37 - Figure 39.

Figure 37 shows that if the probability of experiencing a severe HSR is 0.001 and the dosage of one infusion is 300 ml the cost difference between parenteral and oral iron is CHF 272 per person. For the same probability but a dosage of 500 ml the cost difference is CHF 357 per person and for a dosage of 700 ml CHF 442 per person. The lowest cost difference (CHF 272 per person) between the first-line parenteral and the first-line oral therapy was observed for a dosage of 300 mg per infusion and a probability of severe HSR of 0.1% (light blue line in Figure 37). On the other hand, the highest cost difference (CHF 491 per person) between the first-line parenteral and the first-line oral therapy was observed for a dosage of 700 mg per infusion and a probability of severe HSR of 1.0% (dark blue line in Figure 37).

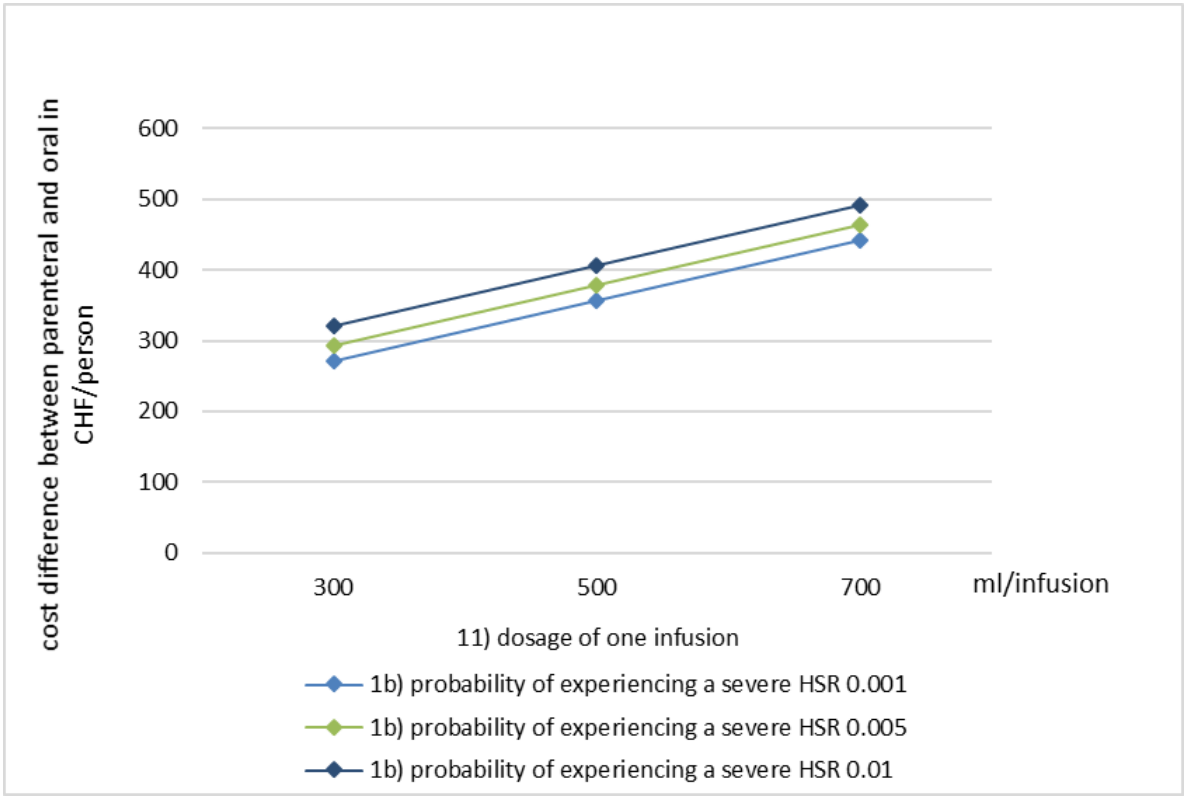


Figure 37 Two-way sensitivity analysis of probability of severe HSR and dosage of parenteral medication

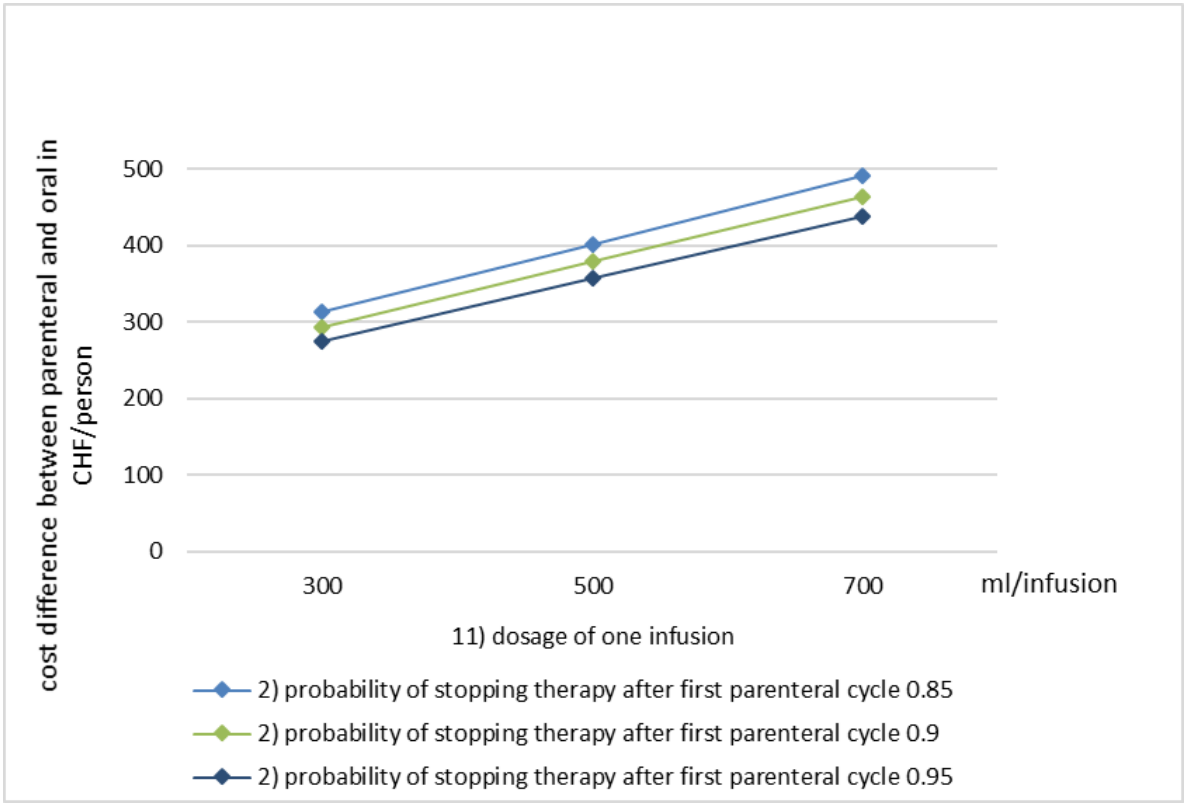


Figure 38 Two-way sensitivity analysis of probability of stopping therapy after first parenteral cycle and dosage of parenteral medication

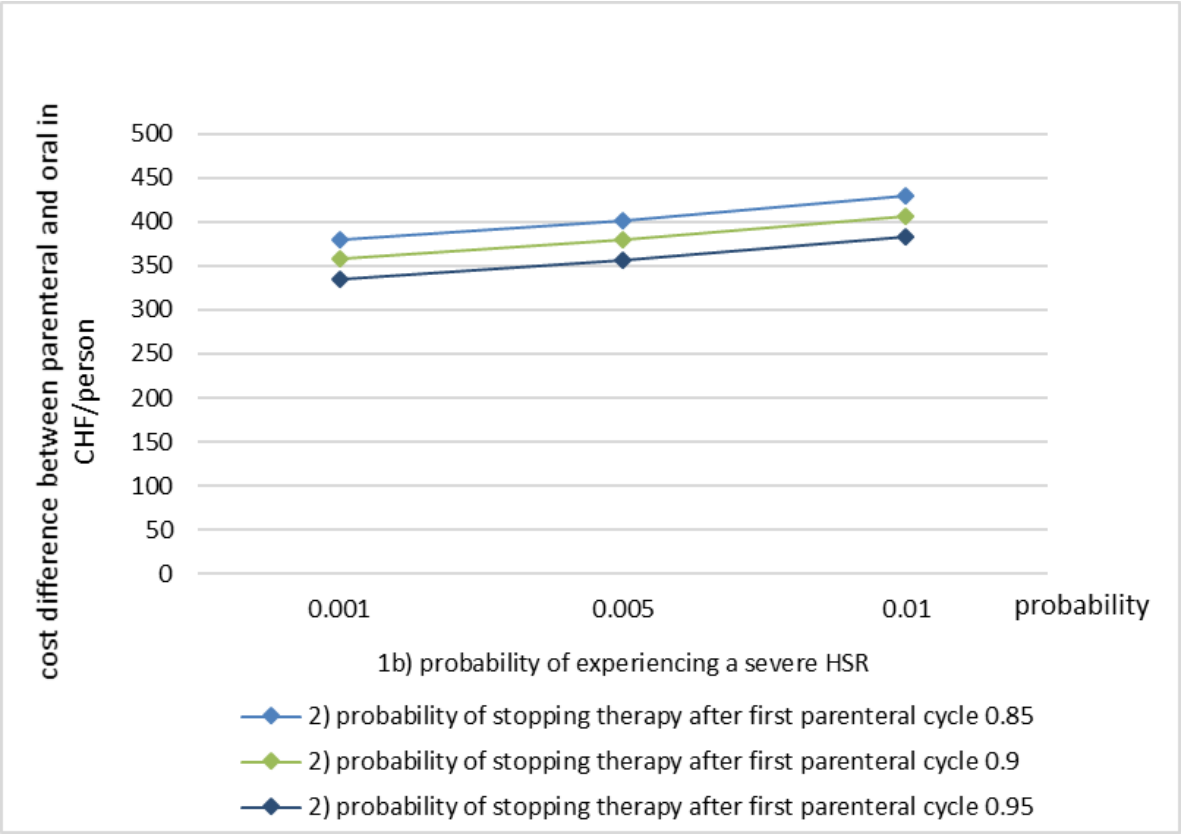


Figure 39 Two-way sensitivity analysis of probability of stopping therapy after first parenteral cycle and probability of a severe HSR

3.3.6 Probabilistic sensitivity analysis

The probabilistic sensitivity analysis (details in section 3.2.4) shows the uncertainty of the point estimates presented as base case results. The estimated cost difference between the two treatment strategies (first-line parenteral and first-line oral iron therapy) varied between CHF 304 and CHF 514 in 95% of all model runs (Table 30). CHF 304 is 20% lower than the result from the base case scenario (CHF 379) and CHF 514 is 36% higher.

Table 30 Probabilistic sensitivity analysis results (in CHF)

	Base case	95% lower bound (Δ%)	95% upper bound
Costs for first-line parenteral	561	471 (-16%)	712 (+27%)
Costs for first-line oral	182	144 (-21%)	224 (+23%)
Cost difference	379	304 (-20%)	514 (+36%)

3.3.7 Budget impact analysis

3.3.7.1 Estimating the target population

Based on information from the Federal Statistical Office regarding the population size older than 18 years and the publication by Biétry et al. (2017), the prevalence of treated iron deficiency patients in Switzerland for 2018 was first estimated by assuming that the patients who received iron therapy were treated for iron deficiency (Table 31)[98].

Table 31 Estimation of the number of patients treated for iron deficiency in Switzerland in 2018

	Estimation	Source
Population ≥18 years	6'963'149	Federal Statistical Office (T 01.02.03.02)
Number of women ≥18 years	3'538'697	Federal Statistical Office (T 01.02.03.02)
Number of men ≥18 years	3'424'452	Federal Statistical Office (T 01.02.03.02)
Prevalence of treated iron deficiency in women	6.3%	Biétry et al. (2017)
Prevalence of treated iron deficiency in men	1.0%	Biétry et al. (2017)
Number of female patients treated for iron deficiency	221'499	
Number of male patients treated for iron deficiency	34'832	
Total number of patients treated for iron deficiency	256'331	

Estimations from the clinical experts involved in this project were then used to calculate the prevalence of treated IDNA patients with fatigue or RLS (Table 32). Two experts felt not comfortable to give any estimations and the estimations given by the other two experts varied widely. Therefore, the mean from both expert opinions was calculated and used in this "mean scenario" as base case target population for the budget impact analysis. The estimation based on expert opinion A served as "lower bound scenario" and the estimation based on expert opinion B as "upper bound scenario".

Table 32 Estimation of the number of treated IDNA patients with fatigue/RLS in Switzerland in 2018

	Estimation		
Number of patients treated for iron deficiency (see Table 31)	256'331		
	Mean of expert opinion A and B	**Expert opinion A**	**Expert opinion B**
Percentage of iron deficiency patients treated for IDNA	51%	25%	78%
Percentage of IDNA patients treated for fatigue/RLS	85%	70%	100%
Number of treated IDNA patients with fatigue/RLS	111'967	44'858	199'368
Percentage of population ≥18 years	1.6%	0.6%	2.9%

Due to internal rounding, the results may differ.

3.3.7.2 Estimating the budget impact

From a health care payer perspective, the costs per patient for first-line parenteral are higher than for first-line oral. Therefore, increasing the use of first-line parenteral always leads to additional costs.

Assuming that in 2018 24.4% instead of 0% of the patients would have been treated with first-line parenteral iron, additional costs of CHF 10.3 million would result from a healthcare payer perspective (Table 33). If the uncertainty regarding the size of the target population is considered, these additional costs are between CHF 4.1-18.4 million. If the uncertainty in the cost difference between the two treatment strategies is also considered, these additional costs are between 3.3-25.0 million.

If a rather hypothetical extreme scenario is assumed, meaning that all patients in 2018 would have been treated with first-line parenteral instead of first-line oral, this would have led to additional costs of CHF 42.4 million. Again, considering the uncertainty in the size of the target population, these additional costs are between CHF 17.0-75.5 million. If the uncertainty in the cost difference between the two treatment strategies is also considered, these additional costs are between CHF 13.6-102.6 million.

Table 33 Budget impact analysis from a health care payer perspective for Switzerland in 2018

Costs per patient (in CHF, see Table 30)	**Base case**	**Lower bound**	**Upper bound**
First-line parenteral	560.75	471.41	712.18
First-line oral	181.91	144.31	224.50
Cost difference	378.84	304.26	514.45
Target population (see Table 32)	**Base case**	**Lower bound**	**Upper bound**
Number of treated IDNA patients with fatigue/RLS	111'967	44'858	199'368
Scenario "base case costs"	**Total costs (in CHF)**		
Share of patients treated with first-line oral (parenteral):	**Base case**	**Lower bound**	**Upper bound**
100% (0%)	20'367'854	8'160'094	36'267'086
76% (24%)	30'717'712	12'306'619	54'696'084
50% (50%)	41'576'578	16'657'071	74'031'427
25% (75%)	52'180'940	20'905'560	92'913'598
0% (100%)	62'785'302	25'154'048	111'795'769
	Budget impact (in CHF)		
	Base case	**Lower bound**	**Upper bound**
costs increasing first-line parenteral from 0% to 24%	10'349'857	4'146'525	18'428'999
costs increasing first-line parenteral from 0% to 100%	42'417'448	16'993'954	75'528'683
costs increasing first-line parenteral by 10%	4'241'745	1'699'395	7'552'868
	Budget impact (in CHF)		
Scenario "minimum cost difference" (in CHF)	**Base case**	**Lower bound**	**Upper bound**
costs increasing first-line parenteral from 0% to 24%	8'312'342	3'330'223	14'800'990
costs increasing first-line parenteral from 0% to 100%	34'066'975	13'648'454	60'659'796
costs increasing first-line parenteral by 10%	3'406'697	1'364'845	6'065'980
	Budget impact (in CHF)		
Scenario "maximum cost difference" (in CHF)	**Base case**	**Lower bound**	**Upper bound**
costs increasing first-line parenteral from 0% to 24%	14'054'704	5'630'819	25'025'864
costs increasing first-line parenteral from 0% to 100%	57'601'246	23'077'129	102'565'017
costs increasing first-line parenteral by 10%	5'760'125	2'307'713	10'256'502

Due to internal rounding, the results may differ.

3.4 Discussion

3.4.1 Summary of the results

A decision tree was built with the aim to reflect the daily practice of general practitioners in Switzerland. Although the model may look sophisticated and many variables had to be parametrized based on expert opinion, the performed model validation showed that 87.5% of the patients with a first-line parenteral treatment strategy and 85.3% of the patients with a first-line oral treatment strategy experience treatment success within the first treatment cycle. These results are consistent with the proportions suggested by the clinical experts who initially assumed between 80% and 90% of the patients are successfully treated within the first cycle. Among both treatment strategies, the probability of a patient to achieve treatment success within the first two treatment cycles (three or six months) amounted to 99%.

Our cost-comparison analysis estimated total direct medical costs from a health care payer perspective for patients with IDNA and fatigue or RLS treated with first-line parenteral iron at CHF 561 per patient and with first-line oral iron at CHF 182 per patient over a time horizon of one year (reference year 2018). The cost difference between the two treatment strategies was estimated at CHF 379 per patient. The univariate sensitivity analysis showed that the following parameters have the largest impact on the result:

- Dosage of the parenteral administration (impact +/-21.2%)
- Duration of visit for a parenteral treatment (impact +14.8%; no lower bound defined)
- Probability of experiencing a severe HSR (impact -5.4%; +6.4%)

The smallest effect on the cost difference was caused by the probability of having a lethal HSR after a parenteral treatment (impact +/-0% compared to the base case difference). In the probabilistic sensitivity analysis, the estimated cost difference between the two treatment strategies (first-line parenteral and first-line oral iron therapy) varied between CHF 304 and CHF 514 in 95% of all model runs, indicating substantial uncertainty.

For the budget impact analysis, it was assumed that 24.4% instead of 0% of the patients would have been treated with first-line parenteral iron in Switzerland in 2018. This led to additional costs of CHF 10.3 million from a health care payer perspective. Considering the uncertainty regarding the size of the target population and the uncertainty in the cost difference between the two treatment strategies, these additional costs were estimated between CHF 3.3-25.0 million. Assuming a rather hypothetical extreme scenario, meaning that all patients in 2018 would have been treated with first-line parenteral instead of first-line oral, this would have led to additional costs of CHF 42.4 million. Considering the uncertainty, these additional costs were estimated between CHF 13.6-102.6 million.

3.4.2 Comparison with existing literature

A previous report by the Swiss Medical Board has estimated direct medical costs of oral versus parenteral iron treatment in patients with iron deficiency (with or without anaemia) from a health care payer perspective[121]. The report assumed that costs for general practitioner visits and labs did not differ between oral and parenteral iron treatment. For the oral iron treatment they estimated costs of approximately CHF 100 based on the assumption that the patients were treated with 200 mg iron daily for 16 weeks. For parenteral iron treatment, they considered costs for the drug, material, venous access and surveillance of the patient. They used a dosage of 1000 mg iron and estimated costs for the

parenteral treatment at approximately CHF 510. The cost difference between parenteral and oral iron was CHF 410. This difference is close to the difference in the present assessment (CHF 379), although higher costs for the two treatment strategies were estimated. For first-line oral iron treatment, a lower dosage was used, but some patients were allowed to take oral iron for up to nine month or to switch to parenteral iron. The costs for parenteral iron were higher in the present calculations because costs due to side effects were included. However, the present model is deemed to better reflect daily practice of general practitioners treating IDNA patients in Switzerland.

For the budget impact analysis, the report by the Swiss Medical Board assumed that 15% of the total population suffer from iron deficiency and that 5% of the patients with iron deficiency suffer from a symptomatic, severe iron deficiency or iron deficiency anaemia and are therefore treated with oral or parenteral iron. This led to 60'000 patients treated with iron. The target population for the budget impact analysis was estimated based on a recent study from Switzerland[98] and expert opinions. The number of treated IDNA patients with fatigue/RLS was estimated at 111'967. The expert opinions varied substantially, therefore a lower bound of the target population (44'858) and an upper bound (199'368) were also estimated.

The Swiss Medical Board estimated additional costs of CHF 25 million assuming that all patients are treated with parenteral instead of oral iron. For such a hypothetical extreme scenario, additional costs of CHF 42.4 million were estimated in the assessment. The differences between the two reports are mainly driven by the different sizes of the target populations.

3.4.3 Strength

To the best of our knowledge, this is the first cost-comparison model developed specifically for patients with IDNA and fatigue or RLS. Based on the model validation, it can be said that the model seems to be representative for the daily practice of general practitioners in Switzerland. In comparison to the report by the Swiss Medical Board, patients were allowed to switch from oral to parenteral and vice versa what represents daily routine in the Swiss setting. Furthermore, it was considered that some patients may need a longer oral treatment than 16 weeks. In regard to parenteral treatment, side effects that are related to substantial costs and the fact that some patients need more than one injection were taken into account. Furthermore, the substantial uncertainty was analysed in univariate, multivariate and probabilistic sensitivity analyses.

3.4.4 Limitations

Substantial uncertainty of the assessment is due to the limited evidence available. Thus, many variables had to be parametrized based on expert opinion and some opinions differed substantially between experts. For some of the variables with available evidence, e.g. probability of experiencing a mild/moderate HSR, the reporting was poor. In addition, other challenges such as the use of different categorizations of side effects were present. However, the univariate sensitivity analysis showed that some parameters with high uncertainty, such as the probability of experiencing a mild/moderate HSR, do not have a relevant influence on the results.

The budget impact analysis was based on recent evidence available in Switzerland. However, the report by Biétry et al. (2017) used claims data for the analysis and did not identify patients who used over-the-counter oral iron therapy (without a prescription)[98]. Consequently, it may underestimate the prevalence of iron therapy. Furthermore, the prevalence available for women and men for all age groups was applied to the population older than 18 years. As the prevalence in patients below the age

of 18 years is smaller, the prevalence in the population older than 18 years may be further underestimated. In addition, it is not known how representative patients insured by Helsana are for the general Swiss population. Moreover, IDNA patients may experience a relapse after an initial successful iron therapy. Such relapses likely exceed the one-year time-horizon investigated in the current analysis, also no data on relapse rates was available for the two treatment strategies and hence, for these two reasons, relapses were not considered in the present assessment. When relapse rates differ between the two treatment strategies, this may lead to additional cost differences between the two treatment strategies. As a further limitation, future population changes and potential changes in the disease awareness in the future were not considered.

As commissioned by the SFOPH, this study was conducted from a health care payer perspective and did not include productivity losses. However, from a societal perspective, productivity losses may be relevant.

3.5 Conclusion

To the best of our knowledge, the present assessment is the first to estimate the cost difference between a first-line parenteral treatment strategy and a first-line oral treatment strategy for adult patients with IDNA and fatigue or RLS from a healthcare payer perspective in Switzerland. The cost of the first-line parenteral treatment strategy was estimated to be CHF 379 per patient higher than first-line oral (CHF 561 versus CHF 182). The results seem to be plausible compared to previous estimations for patients with IDA or symptomatic, severe iron deficiency. Although the findings in the present assessment are partly in line with a similar report, it was shown that the observed cost difference between first-line parenteral and first-line oral iron therapy are subjected to substantial uncertainty. In the probabilistic sensitivity analysis, the estimated cost difference between the two treatment strategies varied between CHF 304 and CHF 514 in 95% of all model runs. For the budget impact analysis, it was assumed that 24.4% instead of 0% of the patients would have been treated with first-line parenteral iron in Switzerland in 2018 and additional costs of CHF 10.3 million were estimated for such a scenario. Considering the uncertainty regarding the size of the target population and the uncertainty in the cost difference between the two treatment strategies these additional costs were estimated to vary between CHF 3.3-25.0 million. Due to the substantial uncertainty in the results, further research regarding dosage and duration of visit for parenteral treatment, probability of experiencing a severe HSR, the prevalence of IDNA patients with fatigue and RLS and the frequency of parenteral iron therapy as first-line treatment seems to be indicated.

4 Reference list

1. Bundesamt für Gesundheit (BAG). Re-Evaluation - HTA: Iron Therapy for Iron Deficiency without Anemia - Final Scope Report 2015 (14.09.2017). *https://www.bag.admin.ch/dam/bag/en/dokumente/kuv-leistungen/bezeichnung-der-leistungen/Re-Evaluation-HTA/eisentherapie-eisenmangel-ohne-anaemie-bag-scopingbericht-2015.pdf.download.pdf/Eisentherapie%20bei%20Eisenmangel%20ohne%20An%C3%A4mie%20-%20BAG-Scopingbericht%202015.pdf.* 2015.
2. Swiss Federal Office of Public Health. Iron Therapy for Iron Deficiency without Anemia – Scope Effectiveness Assessement 2017. *https://www.bag.admin.ch/dam/bag/de/dokumente/kuv-leistungen/bezeichnung-der-leistungen/Re-Evaluation-HTA/iron-therapy--iron-deficiency-without-anemia-final-scope-2017%20.pdf.* 2017.
3. Mattli RR, M.; Glinz, D.; Raatz, H.; Bucher, H. C.; Wieser, S. Economic evaluation of oral versus parenteral iron therapy for iron deficiency without anemia. Scope. . *https://www.bag.admin.ch/dam/bag/de/dokumente/kuv-leistungen/bezeichnung-der-leistungen/Re-Evaluation-HTA/-irontherapy-scope2018.pdf.* 2018.
4. Fehr, Krayenbühl, Favrat, Schleiffenbaum, Kapanci. Diagnose und Behandlung von Eisenmangel ohne Anämie. *Praxis.* 2009;98(24):1445-1451.
5. Herklotz RH, Andreas. Labordiagnose von Eisenstoffwechselstörungen. *Schweiz Med Forum.* 2010;10((30-31)):500–507.
6. Martius F. Eisenmangel ohne Anämie – ein heisses Eisen?: Nichthämatologische Auswirkungen des Eisenmangels: Welche sind belegt, wann kommen sie zum Tragen. *Swiss Medical Forum.* 2009;9(15-16):294-299.
7. WHO. *Assessing the Iron Status of populations. Second edition. Report of a Joint World Health Organization/Centers for Disease Control and Prevention Technical Consultation on the Assessment of Iron Status at the Population Level.* Geneva, Switzerland: World Health Organization; 2007.
8. Guyatt GH, Oxman AD, Ali M, Willan A, McIlroy W, Patterson C. Laboratory diagnosis of iron-deficiency anemia: an overview. *J Gen Intern Med.* 1992;7(2):145-153.
9. Verdon F, Burnand B, Stubi CL, et al. Iron supplementation for unexplained fatigue in non-anaemic women: double blind randomised placebo controlled trial. *BMJ.* 2003;326(7399):1124.
10. Clenin GE. The treatment of iron deficiency without anaemia (in otherwise healthy persons). *Swiss Med Wkly.* 2017;147:w14434.
11. Guyatt G, Oxman AD, Akl E, et al. GRADE guidelines 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol.* 2010.
12. Guyatt G, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol.* 2010.
13. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol.* 2011.
14. Guyatt G, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence-study limitations (risk of bias) and publication bias. *J Clin Epidemiol.* 2011.
15. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. *J Clin Epidemiol.* 2011;64(12):1277-1282.
16. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. *J Clin Epidemiol.* 2011;64(12):1283-1293.
17. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. *J Clin Epidemiol.* 2011;64(12):1294-1302.
18. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. *J Clin Epidemiol.* 2011;64(12):1303-1310.

19. Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J.Clin.Epidemiol.* 2011;64(12):1311-1316.
20. Brunetti M, Shemilt I, Pregno S, et al. Grade guidelines: 10. Considering resource use and rating the quality of economic evidence. *J Clin Epidemiol.* 2012.
21. Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol.* 2012.
22. Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines 12. Preparing Summary of Findings tables-binary outcomes. *J Clin Epidemiol.* 2012.
23. Guyatt GH, Thorlund K, Oxman AD, et al. GRADE guidelines: 13. Preparing Summary of Findings tables and evidence profiles - Continuous outcomes. *Journal of Clinical Epidemiology.* 2013;66(2):173-183.
24. Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol.* 2013;66(7):719-725.
25. Andrews JC, Schunemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol.* 2013;66(7):726-735.
26. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology. *Journal of Clinical Epidemiology.* 2011;64(4):380-382.
27. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions.* Vol Version 5.1.0. www.cochrane.org/training/cochrane-handbook The Cochrane Collaboration last edited 20 March 2011
28. Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. *J Med Libr Assoc.* 2006;94(1):41-47.
29. GRADEpro GDT: GRADEpro Guideline Development Tool [Software]. McMaster University, 2015 (developed by Evidence Prime, Inc.). Available from gradepro.org.
30. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986;7(3):177-188.
31. Cohen J. *Statistical power analysis for the behavioral sciences. 2nd edition. Hillsdale, NJ: Lawrence Earlbaum Associates.* 1988.
32. Jüni P, Reichenbach S, Dieppe P. Osteoarthritis: rational approach to treating the individual. *Best Pract Res Clin Rheumatol.* 2006;20(4):721-740.
33. Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I(2) in assessing heterogeneity may mislead. *BMC Med Res Methodol.* 2008;8:79.
34. Phillips RS, Sung L, Ammann RA, et al. Predicting microbiologically defined infection in febrile neutropenic episodes in children: global individual participant data multivariable meta-analysis. *Br J Cancer.* 2016;114(6):623-630.
35. Walwyn R, Roberts C. Meta-analysis of standardised mean differences from randomised trials with treatment-related clustering associated with care providers. *Stat Med.* 2017;36(7):1043-1067.
36. Allen RP, Picchietti D, Hening WA, Trenkwalder C, Walters AS, Montplaisi J. Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. A report from the restless legs syndrome diagnosis and epidemiology workshop at the National Institutes of Health. *Sleep Med.* 2003;4(2):101-119.
37. Allen RP, Adler CH, Du W, Butcher A, Bregman DB, Earley CJ. Clinical efficacy and safety of IV ferric carboxymaltose (FCM) treatment of RLS: a multi-centred, placebo-controlled preliminary clinical trial. *Sleep Medicine.* 2011;12(9):906-913.
38. Cho Y, Allen RP, Earley CJ. Clinical efficacy of ferric carboxymaltose treatment in patient with restless legs syndrome. *Sleep. Conference: 30th annual meeting of the associated professional sleep societies, LLC, SLEEP 2016. Denver, CO united states. Conference start: 20160611.*

Conference end: 20160615. Conference publication: (var.pagings). 2016;39:A227-a228. http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/758/CN-01266758/frame.html.
39. Cho YW, Allen RP, Earley CJ. Clinical efficacy of ferric carboxymaltose treatment in patients with restless legs syndrome. *Sleep Medicine.* 2016;25:16-23.
40. Davis BJ, Rajput A, Rajput ML, Aul EA, Eichhorn GR. A randomized, double-blind placebo-controlled trial of iron in restless legs syndrome. *European Neurology.* 2000;43(2):70-75.
41. Earley CJ, Horska A, Mohamed MA, Barker PB, Beard JL, Allen RP. A randomized, double-blind, placebo-controlled trial of intravenous iron sucrose in restless legs syndrome. *Sleep Medicine.* 2009;10(2):206-211.
42. Grote L, Leissner L, Hedner J, Ulfberg J. A randomized, double-blind, placebo controlled, multi-center study of intravenous iron sucrose and placebo in the treatment of restless legs syndrome. *Movement Disorders.* 2009;24(10):1445-1452.
43. Lee CS, Lee SD, Kang SH, Park HY, Yoon IY. Comparison of the efficacies of oral iron and pramipexole for the treatment of restless legs syndrome patients with low serum ferritin. *European Journal of Neurology.* 2014;21(2):260-266.
44. Yoon I, Lee C, Lee S, Kang S, Park H. Comparison of efficacy between oral iron and dopamine agonists in the treatment of patients with restless legs syndrome with low-normal serum ferritin. *Sleep.* 2013;36:A247. http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/877/CN-01041877/frame.html.
45. Trenkwalder C, Winkelmann J, Oertel W, Virgin G, Roubert B, Mezzacasa A. Single-dose ferric carboxymaltose for the treatment of restless legs syndrome in iron deficient non-anaemic patients-a randomized, placebo-controlled trial. *Journal of Sleep Research. Conference: 23rd Congress of the European Sleep Research Society, ESRS 2016. Italy. Conference Start: 20160913. Conference End: 20160916.* 2016;25:67-68. http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/330/CN-01213330/frame.html.
46. Trenkwalder C, Winkelmann J, Oertel W, Virgin G, Roubert B, Mezzacasa A. Ferric carboxymaltose in patients with restless legs syndrome and nonanemic iron deficiency: A randomized trial. *Mov Disord.* 2017.
47. Wang J, O'Reilly B, Venkataraman R, Mysliwiec V, Mysliwiec A. Efficacy of oral iron in patients with restless legs syndrome and a low-normal ferritin: A randomized, double-blind, placebo-controlled study. *Sleep Medicine.* 2009;10(9):973-975.
48. Abetz L, Arbuckle R, Allen RP, et al. The reliability, validity and responsiveness of the International Restless Legs Syndrome Study Group rating scale and subscales in a clinical-trial setting. *Sleep Med.* 2006;7(4):340-349.
49. Allen RP, Picchietti DL, Garcia-Borreguero D, et al. Restless legs syndrome/Willis-Ekbom disease diagnostic criteria: updated International Restless Legs Syndrome Study Group (IRLSSG) consensus criteria--history, rationale, description, and significance. *Sleep Med.* 2014;15(8):860-873.
50. Walters AS, LeBrocq C, Dhar A, et al. Validation of the International Restless Legs Syndrome Study Group rating scale for restless legs syndrome. *Sleep Med.* 2003;4(2):121-132.
51. Scholz H, Trenkwalder C, Kohnen R, Riemann D, Kriston L, Hornyak M. Dopamine agonists for restless legs syndrome. *Cochrane Database Syst Rev.* 2011(3):CD006009.
52. Krayenbuehl PA, Battegay E, Breymann C, Furrer J, Schulthess G. Intravenous iron for the treatment of fatigue in nonanemic, premenopausal women with low serum ferritin concentration. *Blood.* 2011;118(12):3222-3227.
53. Favrat B, Balck K, Breymann C, et al. Evaluation of a single dose of ferric carboxymaltose in fatigued, iron-deficient women--PREFER a randomized, placebo-controlled study. *PloS one.* 2014;9(4):e94217. http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/050/CN-01117050/frame.html.
54. Favrat B, Balck K, Gasche C, et al. A single 1000mg iron dose of ferric carboxymaltose improves fatigue in iron deficient, non-anaemic premenopausal women - Results of the randomised, placebo-controlled prefer study. *International journal of gynaecology and obstetrics.*

2012;119:S858-s859. http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/383/CN-01004383/frame.html.
55. Favrat B, Balck K, Gasche C, et al. One 1000 mg iron dose of ferric carboxymaltose improved fatigue in iron-deficient, non-anaemic women in the randomised placebo-controlled study PREFER. *Bjog.* 2012;119:232-233. http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/833/CN-01032833/frame.html.
56. Vaucher P, Druais PL, Waldvogel S, Favrat B. Effect of iron supplementation on fatigue in nonanemic menstruating women with low ferritin: a randomized controlled trial. *CMAJ Canadian Medical Association Journal.* 2012;184(11):1247-1254.
57. Konofal E, Lecendreux M, Deron J, et al. Effects of iron supplementation on attention deficit hyperactivity disorder in children. *Pediatric Neurology.* 2008;38(1):20-26.
58. Anttila S, Persson J, Vareman N, Sahlin NE. Conclusiveness resolves the conflict between quality of evidence and imprecision in GRADE. *J Clin Epidemiol.* 2016;75:1-5.
59. Schünemann HJ. Interpreting GRADE's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? *J Clin Epidemiol.* 2016;75:6-15.
60. Earley CJ, Connor J, Garcia-Borreguero D, et al. Altered brain iron homeostasis and dopaminergic function in Restless Legs Syndrome (Willis-Ekbom Disease). *Sleep Med.* 2014;15(11):1288-1301.
61. Schormair B, Zhao C, Bell S, et al. Identification of novel risk loci for restless legs syndrome in genome-wide association studies in individuals of European ancestry: a meta-analysis. *Lancet Neurol.* 2017;16(11):898-907.
62. Connor JR, Patton SM, Oexle K, Allen RP. Iron and restless legs syndrome: treatment, genetics and pathophysiology. *Sleep Med.* 2017;31:61-70.
63. Garcia-Borreguero D, Silber MH, Winkelman JW, et al. Guidelines for the first-line treatment of restless legs syndrome/Willis-Ekbom disease, prevention and treatment of dopaminergic augmentation: a combined task force of the IRLSSG, EURLSSG, and the RLS-foundation. *Sleep Med.* 2016;21:1-11.
64. Savovic J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med.* 2012;157(6):429-438.
65. Allen RP. Minimal clinically significant change for the International Restless Legs Syndrome Study Group rating scale in clinical trials is a score of 3. *Sleep Med.* 2013;14(11):1229.
66. Winkelman JW, Armstrong MJ, Allen RP, et al. Practice guideline summary: Treatment of restless legs syndrome in adults: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology.* 2016;87(24):2585-2593.
67. Rizzo G, Plazzi G. Neuroimaging Applications in Restless Legs Syndrome. *Int Rev Neurobiol.* 2018;143:31-64.
68. Trotti LM, Bhadriraju S, Becker LA. Iron for restless legs syndrome. *Cochrane Database Syst Rev.* 2012(5):CD007834.
69. Hurrell R, Egli I. Iron bioavailability and dietary reference values. *Am J Clin Nutr.* 2010;91(5):1461S-1467S.
70. Ding J, Fan W, Chen HH, Yan P, Sun SG, Zheng J. Rotigotine in the treatment of primary restless legs syndrome: A meta-analysis of randomized placebo-controlled trials. *J Huazhong Univ Sci Technolog Med Sci.* 2015;35(2):169-175.
71. Liu GJ, Wu L, Lin Wang S, Xu LL, Ying Chang L, Fu Wang Y. Efficacy of Pramipexole for the Treatment of Primary Restless Leg Syndrome: A Systematic Review and Meta-analysis of Randomized Clinical Trials. *Clin Ther.* 2016;38(1):162-179 e166.
72. Zhang W, Wang Y, Cong SY, Nao JF, Feng J, Bi GR. Efficacy and tolerability of pramipexole for the treatment of primary restless leg syndrome: a meta-analysis of randomized placebo-controlled trials. *Neuropsychiatr Dis Treat.* 2013;9:1035-1043.

73. Galan P, Yoon HC, Preziosi P, et al. Determining factors in the iron status of adult women in the SU.VI.MAX study. SUpplementation en VItamines et Mineraux AntioXydants. *Eur J Clin Nutr.* 1998;52(6):383-388.
74. Hercberg S, Preziosi P, Galan P. Iron deficiency in Europe. *Public Health Nutr.* 2001;4(2B):537-545.
75. Yokoi K, Konomi A. Iron deficiency without anaemia is a potential cause of fatigue: meta-analyses of randomised controlled trials and cross-sectional studies. *Br J Nutr.* 2017;117(10):1422-1431.
76. Stockings E, Degenhardt L, Lee YY, et al. Symptom screening scales for detecting major depressive disorder in children and adolescents: a systematic review and meta-analysis of reliability, validity and diagnostic utility. *J Affect Disord.* 2015;174:447-463.
77. *Depression: The Treatment and Management of Depression in Adults (Updated Edition).* Leicester UK: The British Psychological Society & The Royal College of Psychiatrists, 2010.; 2010.
78. Myers K, Winters NC. Ten-year review of rating scales. II: Scales for internalizing disorders. *J Am Acad Child Adolesc Psychiatry.* 2002;41(6):634-659.
79. Auerbach M, Macdougall IC. Safety of intravenous iron formulations: facts and folklore. *Blood Transfus.* 2014;12(3):296-300.
80. Low MS, Speedy J, Styles CE, De-Regil LM, Pasricha SR. Daily iron supplementation for improving anaemia, iron status and health in menstruating women. *Cochrane Database Syst Rev.* 2016;4:CD009747.
81. Beutler E, Larsh SE, Gurney CW. Iron therapy in chronically fatigued, nonanemic women: a double-blind study. *Ann Intern Med.* 1960;52:378-394.
82. Morrow JJ, Dagg JH, Goldberg A. A controlled trial of iron therapy in sideropenia. *Scott Med J.* 1968;13(3):79-83.
83. Waldvogel S, Pedrazzini B, Vaucher P, et al. Clinical evaluation of iron treatment efficiency among non-anemic but iron-deficient female blood donors: a randomized controlled trial. *BMC Med.* 2012;10:8.
84. Houston BL, Hurrie D, Graham J, et al. Efficacy of iron supplementation on fatigue and physical capacity in non-anaemic iron-deficient adults: a systematic review of randomised controlled trials. *BMJ Open.* 2018;8(4):e019240.
85. Millichap JG. Etiologic classification of attention-deficit/hyperactivity disorder. *Pediatrics.* 2008;121(2):e358-365.
86. Pliszka S. Practice parameter for the assessment and treatment of children and adolescents with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry.* 2007;46(7):894-921.
87. Banerjee TD, Middleton F, Faraone SV. Environmental risk factors for attention-deficit hyperactivity disorder. *Acta Paediatr.* 2007;96(9):1269-1274.
88. Mcgoey K, Dupaul G, Haley E, Shelton T. Parent and teacher ratings of attention-deficit/hyperactivity disorder in preschool: the ADHD Rating Scale-IV Preschool Version. *Journal of Psychopathology and Behavioral Assessment* 2007;29(4):269-276.
89. Wolraich M, Brown L, Brown RT, et al. ADHD: clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents. *Pediatrics.* 2011;128(5):1007-1022.
90. Cortese S, Angriman M, Lecendreux M, Konofal E. Iron and attention deficit/hyperactivity disorder: What is the empirical evidence so far? A systematic review of the literature. *Expert Rev Neurother.* 2012;12(10):1227-1240.
91. Hariri M, Azadbakht L. Magnesium, Iron, and Zinc Supplementation for the Treatment of Attention Deficit Hyperactivity Disorder: A Systematic Review on the Recent Literature. *Int J Prev Med.* 2015;6:83.
92. Sever Y, Ashkenazi A, Tyano S, Weizman A. Iron treatment in children with attention deficit hyperactivity disorder. A preliminary report. *Neuropsychobiology.* 1997;35(4):178-180.

93. Silva MA, Duarte GS, Camara R, et al. Placebo and nocebo responses in restless legs syndrome: A systematic review and meta-analysis. *Neurology.* 2017;88(23):2216-2224.
94. Trotti LM, Becker LA. Iron for the treatment of restless legs syndrome. *Cochrane Database Syst Rev.* 2019;1:CD007834.
95. Shepshelovich D, Rozen-Zvi B, Avni T, Gafter U, Gafter-Gvili A. Intravenous Versus Oral Iron Supplementation for the Treatment of Anemia in CKD: An Updated Systematic Review and Meta-analysis. *Am J Kidney Dis.* 2016;68(5):677-690.
96. Rampton D, Folkersen J, Fishbane S, et al. Hypersensitivity reactions to intravenous iron: guidance for risk minimization and management. *Haematologica.* 2014;99(11):1671-1676.
97. Swiss Federal Office of Public Health. http://www.spezialitätenliste.ch/. Accessed 22.8.2018.
98. Bietry FA, Hug B, Reich O, Susan JS, Meier CR. Iron supplementation in Switzerland - A bi-national, descriptive and observational study. *Swiss Med Wkly.* 2017;147:w14444.
99. Swiss Medical Association F. https://www.fmh.ch/ambulante_tarife/tarmed-tarif/tarmed_tarifbrowser-datenbank.html. Accessed 22.08.2018.
100. Ingenieurbuero S Moeckli. https://eligo.ch/Tarmed-Taxpunktwerte.html. Accessed 22.08.2018.
101. Swiss Federal Office of Public Health. https://www.bag.admin.ch/bag/de/home/themen/versicherungen/krankenversicherung/krankenversicherung-leistungen-tarife/Mittel-und-Gegenstaendeliste.html. Accessed 22.08.2018.
102. Wieser S, Ruthemann I, De Boni S, et al. Cost of acute coronary syndrome in Switzerland in 2008. *Swiss Med Wkly.* 2012;142:w13655.
103. Bundesamt für Statistik (BFS). https://www.bfs.admin.ch/bfs/de/home/statistiken/preise/landesindex-konsumentenpreise/lik-resultate.assetdetail.5866100.html. Accessed 22.08.2018.
104. Swiss DRG AG. https://grouper.swissdrg.org/swissdrg/single. Accessed 22.08.2018.
105. Schneider R, Schur N, Reinau D, Schwenkglenks M, Meier CR. *Helsana-Arzneimittelreport für die Schweiz 2017.* Zürich: Helsana;2017.
106. Briggs AH, Claxton K, Sculpher MJ. *Decision modelling for health economic evaluation.* Oxford university press; 2006.
107. Sullivan SD, Mauskopf JA, Augustovski F, et al. Budget impact analysis-principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. *Value Health.* 2014;17(1):5-14.
108. Andersson M, Egli IM, Zimmermann MB. Eisenmangel. 2011;ARS MEDICI DOSSIER(VII):20-25.
109. Hug BL, Tichelli A, Benkert P, Stirnimann G, Schifferli JA. Diagnosis and treatment of iron deficiency in medical inpatients at a Swiss tertiary university referral hospital: a retrospective observational cohort study of clinical practice. *Swiss Med Wkly.* 2013;143:w13847.
110. Schleiffenbaum BE, Schaer DJ, Burki D, et al. Unexpected high prevalence of metabolic disorders and chronic disease among young male draftees--the Swiss Army XXI experience. *Swiss Med Wkly.* 2006;136(11-12):175-184.
111. Schuepbach RA, Bestmann L, Bechir M, Fehr J, Bachli EB. High Prevalence of Iron Deficiency among Educated Hospital Employees in Switzerland. *Int J Biomed Sci.* 2011;7(2):150-157.
112. Bundesamt für Statistik (BFS). https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/tabellen.assetdetail.5866882.html. Accessed 23.10.2018.
113. swissmedic. http://www.swissmedicinfo.ch/. Accessed 22.08.2018.
114. Broche D-E, Gay C, Armand-Branger S, Grangeasse L, Terzibachian J-J. Severe anaemia in the immediate post-partum period. Clinical practice and value of intravenous iron. *European Journal of Obstetrics & Gynecology and Reproductive Biology.* 2005;123:S21-S27.
115. Quintana-Diaz M, Munoz-Romo R, Gomez-Ramirez S, et al. A fast-track anaemia clinic in the Emergency Department: cost-analysis of intravenous iron administration for treating iron-deficiency anaemia. *Blood Transfus.* 2017;15(5):438-446.

116. Diez‐Lobo AI, Fisac‐Martín MP, Bermejo‐Aycar I, MuÑoz M. Preoperative intravenous iron administration corrects anemia and reduces transfusion requirement in women undergoing abdominal hysterectomy. *Transfusion alternatives in transfusion medicine.* 2007;9(2):114-119.
117. Chertow GM, Mason PD, Vaage-Nilsen O, Ahlmen J. Update on adverse drug events associated with parenteral iron. *Nephrol Dial Transplant.* 2006;21(2):378-382.
118. Suominen P, Punnonen K, Rajamaki A, Irjala K. Serum transferrin receptor and transferrin receptor-ferritin index identify healthy subjects with subclinical iron deficits. *Blood.* 1998;92(8):2934-2939.
119. Zaim M, Piselli L, Fioravanti P, Kanony-Truc C. Efficacy and tolerability of a prolonged release ferrous sulphate formulation in iron deficiency anaemia: a non-inferiority controlled trial. *Eur J Nutr.* 2012;51(2):221-229.
120. Paesano R, Berlutti F, Pietropaoli M, Goolsbee W, Pacifici E, Valenti P. Lactoferrin efficacy versus ferrous sulfate in curing iron disorders in pregnant and non-pregnant women. *Int J Immunopathol Pharmacol.* 2010;23(2):577-587.
121. Swiss Medical Board. Assessment Bericht: Orale oder parenterale Behandlung des Eisenmangels. www.medical-board.ch/fileadmin/docs/public/mb/fachberichte/2014-10-24_eisenmangelbericht_def.pdf. 2014.
122. Biétry F, Schur N, Pfeil A, Schwenkglenks M, Meier CR. Helsana-Arzneimittelreport. Ausgabe 2015. https://www.helsana.ch/docs/arzneimittelreport-2015.pdf. 2015.
123. Evstatiev R, Marteau P, Iqbal T, et al. FERGIcor, a randomized controlled trial on ferric carboxymaltose for iron deficiency anemia in inflammatory bowel disease. *Gastroenterology.* 2011;141(3):846-853.e841-842.
124. Malone M, Barish C, He A, Bregman D. Comparative review of the safety and efficacy of ferric carboxymaltose versus standard medical care for the treatment of iron deficiency anemia in bariatric and gastric surgery patients. *Obes Surg.* 2013;23(9):1413-1420.
125. Chertow GM, Winkelmayer WC. On the relative safety of intravenous iron formulations: new answers, new questions. *American journal of hematology.* 2010;85(9):643-644.
126. Wysowski DK, Swartz L, Borders-Hemphill BV, Goulding MR, Dormitzer C. Use of parenteral iron products and serious anaphylactic-type reactions. *Am J Hematol.* 2010;85(9):650-654.
127. Bailie GR, Mason NA, Valaoras TG. Safety and tolerability of intravenous ferric carboxymaltose in patients with iron deficiency anemia. *Hemodial Int.* 2010;14(1):47-54.
128. Wetmore JB, Weinhandl ED, Zhou J, Gilbertson DT. Relative Incidence of Acute Adverse Events with Ferumoxytol Compared to Other Intravenous Iron Compounds: A Matched Cohort Study. *PLoS One.* 2017;12(1):e0171098.
129. Patterson AJ, Brown WJ, Roberts DC, Seldon MR. Dietary treatment of iron deficiency in women of childbearing age. *Am J Clin Nutr.* 2001;74(5):650-656.
130. Leonard AJ, Chalmers KA, Collins CE, Patterson AJ. Comparison of two doses of elemental iron in the treatment of latent iron deficiency: efficacy, side effects and blinding capabilities. *Nutrients.* 2014;6(4):1394-1405.

5 Appendices

5.1 Appendix – Search strategy for Medline OvidSP and CENTRAL

Appendix 1 Search strategy for Medline and Central

5.1.1 Medline via OvidSP

Datenbank: Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R)

Suchstrategie:

1	ferrous.ti,ab.	(10920)
2	ferric.ti,ab.	(15885)
3	iron.ti,ab.	(153266)
4	1 or 2 or 3	(165203)
5	exp Iron/ad, tu, th [Administration & Dosage, Therapeutic Use, Therapy]	(7384)
6	exp Iron Compounds/ad, tu, th [Administration & Dosage, Therapeutic Use, Therapy]	(8547)
7	exp iron, dietary/	(2654)
8	4 or 5 or 6 or 7	(170241)
9	therapy.ti,ab.	(1496713)
10	administration.ti,ab.	(721385)
11	intake.ti,ab.	(220720)
12	supplement*.ti,ab.	(263116)
13	replac*.ti,ab.	(357190)
14	therapeutic.ti,ab.	(778397)
15	administered.ti,ab.	(469099)
16	exp therapeutics/	(3824231)
17	treat*.ti,ab.	(4536845)
18	exp Dietary Supplements/	(56559)
19	exp Pharmaceutical Preparations/th [Therapy]	(248)
20	9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19	(8623123)
21	gluconate.ti,ab.	(6269)
22	sucrose.ti,ab.	(58069)
23	dextran.ti,ab.	(30860)
24	carboxymaltose.ti,ab.	(250)
25	isomaltoside.ti,ab.	(85)
26	ferumoxytol.ti,ab.	(249)
27	21 or 22 or 23 or 24 or 25 or 26	(94291)
28	sulphate.ti,ab.	(32120)
29	sulfate.ti,ab.	(124510)
30	gluconate.ti,ab.	(6269)
31	lactate.ti,ab.	(88604)
32	bisglycinate.ti,ab.	(28)
33	citrate.ti,ab.	(37227)
34	edta.ti,ab.	(32508)
35	fumarate.ti,ab.	(7286)
36	succinate.ti,ab.	(20389)
37	saccharate.ti,ab.	(133)
38	orthophospate.ti,ab.	(3)
39	pyrophosphate.ti,ab.	(13572)
40	electrolytic.ti,ab.	(5855)
41	28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40	(351956)

42	randomized controlled trial.pt.	(448956)
43	controlled clinical trial.pt.	(91953)
44	randomized.ab.	(389662)
45	randomised.ab.	(77010)
46	placebo.ab.	(184067)
47	clinical trials as topic.sh.	(181513)
48	randomly.ab.	(272044)
49	Random*.tw.	(916326)
50	trial.ti.	(174720)
51	42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50	(1321316)
52	exp animals/ not humans.sh.	(4326005)
53	51 not 52	(1210682)
54	8 and (20 or 27 or 41) and 53	(5631)

5.1.2 CENTRAL

ID	Search	Hits
#1	iron:ti,ab,kw	6107
#2	ferrous:ti,ab,kw	985
#3	ferric:ti,ab,kw	835
#4	{or #1-#3}	6470
#5	MeSH descriptor: [Iron] explode all trees	1836
#6	#1 or #2 or #3 or #5	6470
#7	therapy:ti,ab,kw	319879
#8	administration:ti,ab,kw	176607
#9	intake:ti,ab,kw	30539
#10	supplement*:ti,ab,kw	39825
#11	replac*:ti,ab,kw	24595
#12	therapeutic:ti,ab,kw	57085
#13	administered:ti,ab,kw	70573
#14	treat*:ti,ab,kw	516792
#15	MeSH descriptor: [Therapeutics] explode all trees	280188
#16	MeSH descriptor: [Dietary Supplements] explode all trees	9665
#17	MeSH descriptor: [Pharmaceutical Preparations] explode all trees	63633
#18	{or #7-#17}	758754
#19	(gluconate or sucrose or dextran or carboxymaltose or isomaltoside or ferumoxytol):ti,ab,kw	4213
#20	(sulphate or sulfate or gluconate or lactate or bisglycinate or citrate or edta or fumarate or succinate or saccharate or orthophospate or pyrophosphate or electrolytic):ti,ab,kw	23720
#21	#6 and (#18 or #19 or #20)	5543

5.2 Appendix – Eligibility criteria

Appendix 2 Eligibility criteria of the included RCTs

Study	Inclusion criteria	Exclusion criteria
Allen 2011	"Patients at least 18 years old diagnosed at the clinical centre with RLS based on the IRLS diagnostic criteria were included if they were able to give informed consent after they read and signed the consent form approved by the enrolling institution. They had to have regular sleep hours between 21:00 and 09:00, an IRLS baseline score ≥15, RLS symptoms occurring ≥5 nights per week, an actigraph measured PLMS (PAM-RL) average for 3–5 nights ≥15 h^{-1}. Their RLS diagnosis was independently confirmed by use of the validated Hopkins Telephone Diagnostic Interview conducted by an RLS expert trained in the use of this instrument. Subjects also had to discontinue any use of anti-depressants, sleep medications, dopamine agonists, benzodiazepines, narcotics, or other RLS treatments for at least one week or five half-lives, whichever was longer, before any baseline RLS assessments and PLMS measurements were obtained (non-narcotic analgesics were permitted)."	"Patients were excluded from the study if they were not practising an acceptable form of birth control while at risk for pregnancy or had RLS secondary to: central nervous system (CNS) disease, CNS injury, or chronic kidney disease. They were also excluded if they had any pain or sleep disorders that would disturb clinical sleep measures or had any disease that would disrupt iron status or evaluations in this study. They were excluded if their at baseline serum ferritin was >300 mcg l1, their TSATP45%, their haemoglobin> normal, or if they had other abnormal clinical evaluations. (Online Supplementary data lists all exclusion criteria for the study.)"
Cho 2016*	"Primary RLS patients >18 years of age who had no co-morbid medical disease were enrolled. The diagnosis of RLS was established by a neurologist using the Korean version of the Hopkins–Hening Telephone Diagnostic questionnaire (HTDQ) during a face-to-face interview which conforms to the updated International Restless Legs Syndrome Study Group (IRLSSG) diagnostic criteria. Any patient whose symptoms occurred more than five nights per week and had a score on the International RLS Severity scale (IRLSS scale) of ≥15 when off of all RLS medications for at least 14 days were eligible for enrollment in this study. Treatments for RLS (if any), including antidepressants, hypnotics, dopamine agonists, benzodiazepines, and narcotics, were stopped at least two weeks before baseline assessments."	"Exclusion criteria for this study were as follows: secondary RLS (due to polyneuropathy, neurodegenerative disease, chronic kidney disease, pregnancy), medications that have an influence on RLS symptoms that could not be stopped (eg, antipsychotics and antidepressants), history of hypersensitivity to i.v. iron, severe medical diseases that could disturb iron metabolism or could not withstand FCM (eg, chronic liver disease, chronic heart failure, chronic renal failure), serum ferritin >300 ng/dL, serum hemoglobin <12 g/dL, or transferrin saturation ≥45%."
Davis 2000	"To be included in the study, patients had to have symptomatic RLS and be under treatment at the time of enrollment. […] Patients were included regardless of other potential causes of RLS, such as neuropathy, renal disease, etc."	"Exclusion criteria included allergy to iron sulfate, anemia (hemoglobin <10), current or recent treatment with iron sulfate (200 mg or more per day for at least half of the days in the past 6 months), current pregnancy, hemochromatosis, peptic ulcer disease, history of gastrointestinal neoplasm within the past 2 years, active bacterial infection, or

Study	Inclusion criteria	Exclusion criteria
		current treatment with medications known by the patients to exacerbate their RLS."
Earley 2009	"Following the serology assessment a potential candidate was then evaluated using the Johns Hopkins telephone diagnostic interview with an RLS expert (RPA) conducting the interview. […] All subjects were required to stop consuming any herbal agents or over-the-counter vitamins which might contain iron at least one week prior to the treatment initiation and were required not to use any of these supplements until the conclusion of the study. Any medications that were being used to treat the RLS symptoms were discontinued at least one week prior to the GCRC visit. The patient was required to cease using all other treatments for their RLS for the duration of the study. Patients were instructed not to donate blood for at least 6 weeks prior to the study and not to donate blood as long as they remained in the study."	"Exclusion criteria included: possible secondary forms of RLS; hemoglobin <12 g/dl; any pain-related conditions or any other sleep related problems that might interfere with the interpretation of the outcome measures; sleep apnea rates>25/h; any organ problems (by history or blood study), that would affect RLS symptoms or the treatment with iron. Patients were required to have periodic leg movements of sleep (PLMS), >15/h on the second-night polysomnogram, which was performed during their stay in the General Clinical Research Center (GCRC)."
Grote 2009	"Criteria for inclusion were age between 18 and 70 years, 4 cardinal RLS diagnostic criteria,20 a score of 10 or more on the International Restless Legs Study Group Rating Scale (IRLS), a S-ferritin concentration below 30 lg/L and normal folic acid/ B12 vitamin serum values (Table 1). A study amendment issued after inclusion of 30 patients increased the threshold for S-ferritin to 45 lg/L according to previously published recommendations."	"Exclusion criteria encompassed concomitant use of any drug treatment for RLS, clinical or laboratory findings suggestive of secondary RLS, any previously known clinically significant allergic reaction, use of drug treatment known to induce RLS, pregnancy or a specific contraindication for iron sucrose."
Lee 2014	"Criteria for inclusion were a diagnosis of RLS, age between 20 and 80 years, and a serum ferritin concentration between 15 and 50 ng/ml. Diagnoses were established by face-to-face interview with two psychiatrists specializing in sleep disorders using the diagnostic criteria for RLS recommended by the National Institutes of Health."	"Subjects who were pregnant and those with a history of hemochromatosis, severe liver disease, end-stage renal disease or malignancy were excluded. In addition, subjects allergic to iron were excluded and those who had been on iron replacement or medication affecting RLS symptoms, such as antidepressants, antipsychotics, anticonvulsants, anxiolytics or hypnotics, during the previous 2 months."
Trenkwalder 2017	"Patients aged >18 years weighing >50 kg with moderate to severe RLS (International RLS Severity Scale [IRLS] total score ≥15), normal hemoglobin levels (women, >11.5 g/dL; men, >12.5 g/dL), and serum ferritin <75 lg/L were eligible for this study (patients were also included if serum ferritin was between 75 and 300 lg/L and transferrin saturation [TSAT] was <20%)." "The inclusion criteria also specified patients either to be naïve to RLS medication or not to have taken any RLS medication for at least 7 days prior to study initiation." (Online Supplement)	"Patients were excluded if they had a history or presence of severe psychiatric disorder, history of severe systemic diseases or clinically relevant hepatic dysfunction, current augmentation of restless leg syndrome (RLS), acute or chronic infection, known relevant cardiac dysfunction and/or arrhythmias, known history or presence of moderate/severe pain disorders, hemoglobinopathy, hemochromatosis, or other iron-storage disorders." (Online Supplement)

Study	Inclusion criteria	Exclusion criteria
Wang 2009	"Patients gave written consent to be contacted if they met NIH diagnostic criteria for RLS (Table 1), and received a score of P11 using the validated IRLS. These patients were further screened by measuring random levels of hemoglobin, ferritin, iron, and iron saturation percentage. Only those patients with a measured ferritin level of 15–75 ng/ml were included in the study."	"Patients were excluded from the study for pregnancy, hemochromatosis or other significant liver disease, end-stage renal disease, significant sleep disturbances for reasons other than RLS (i.e., known obstructive sleep apnea, periodic limb movements of sleep, etc.), iron saturation less than 15%, hemoglobin levels less than 11.1 g/dL for females and 14 g/dL for males, iron sulfate allergy, current or recent treatment with iron sulfate as defined by more than 325 mg each day for at least half of the days in the past 2 months or any other potential medications for treatment of RLS.
FERRIM	"Premenopausal, menstruating women ≥ 18 years of age who presented with fatigue were evaluated for inclusion in the study. Inclusion criteria were serum ferritin concentration ≤ 50 ng/mL, hemoglobin concentration ≥ 120 g/L, and adequate contraception for the study period."	"Exclusion criteria were pregnancy, intake of gestagens repressing menstruation, physical or mental disorders, medication affecting physical or mental performance, iron treatment in the 4 weeks before enrollment, and history of hypersensitivity to any iron medication."
PREFER	"Eligible patients were premenopausal, regularly menstruating women ≥18 years of age with symptomatic fatigue (≥5 points on the PFS), who had ID with an unknown etiology (e.g., no menorrhagia) but had normal or borderline hemoglobin (Hb ≥ 115 g/L) at screening. Based on recommendations in other indications and similar to the FERRIM study, ID was defined as serum ferritin <50 µg/L and transferrin saturation (TSAT) <20%, or ferritin <15 µg/L. Further inclusion criteria were a body weight of 50–90 kg (to exclude potential overweightrelated impairment of iron metabolism), a negative pregnancy test and normal levels of C-reactive protein, thyroid-stimulating hormone, vitamin B12 and folic acid (according to each centers protocol)."	"Patients were excluded if they had any active or unstable concurrent medical condition, any major depressive disorder, ongoing infections or chronic inflammatory disease, any history of sleep apnea or concurrent medications that could affect physical or mental performance, a known sensitivity to any iron preparation, or use of iron preparations within 4 weeks prior screening."
Vaucher 2012	"To be eligible, the following criteria had to be met: (a) be menstruating women, (b) be between 18 and 50 years old, (c) report considerable fatigue (> 6 on a 1–10 Likert scale) without obvious clinical causes, (d) not have anemia (hemoglobin ≥ 12.0 g/dL), (e) have a low or borderline ferritin level (< 50 µg/L), (f) not have a known pathology that could explain the fatigue (e.g., psychiatric, thyroid, liver, rheumatic, renal, cardiovascular, pulmonary or oncologic cause), (g) not be pregnant or breastfeeding, (h) not have a digestive disorder that could alter the absorption of the study treatment and (i) not already be taking iron supplementation."	n.r.

Study	Inclusion criteria	Exclusion criteria
Verdon 2003	"Women aged 18 to 55 were included if their main reason for consulting was fatigue."	"We excluded women with anaemia (haemoglobin concentration < 117 g/l), other obvious physical or psychiatric cause for fatigue, or chronic fatigue syndrome."
Konofal 2008	"Subjects were outpatient children with attention deficit hyperactivity disorder aged 5-8 years who met DSM-IV diagnostic criteria for attention deficit hyperactivity disorder by clinical assessment and had serum ferritin levels <30 ng/mL (retaining the definition of iron deficiency from a previous study) with normal hemoglobin levels at the screening."	"We excluded potential subjects if they had an IQ < 80 by the French version of the Wechsler Intelligence Scale, third edition, for children, relevant psychiatric comorbidities (depressive, anxiety, and sleep disorders according to DSM-IV criteria), or chronic medical conditions (including malnutrition). We also excluded children who had received iron supplementation in the past 3 months or previous treatment with psychotropic agents or psychostimulants."

*In Cho 2016, an exclusion criterion for serum haemoglobin concentration of <12 µg/dl was reported; however, reviewers came to the conclusion that this was a typographical based on the author's statement of a non-anaemic population error. Therefore, the exclusion criterion for serum haemoglobin was changed from <12 µg/dl to <12 g/dl.

Abbreviations: n.r., not reported

5.3 Appendix – Risk of bias with support for judgement

Appendix 3 Risk of Bias with support for judgement

Study	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete continuous outcome data (attrition bias) and support for judgement	Incomplete binary data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
Adults with Restless legs syndrome							
Allen 2011	Unclear n.r.	Low "double-blinded procedures with the randomisation managed and recorded at a central location not at the study sites. […] All subjects, investigators, and study personnel were blinded to the content of the study drug, with the exception of the unblinded study personnel (at most sites a study nurse and a back-up study nurse) who were responsible for the following: Randomising the subject on day 0 through the use of an interactive voice recognition system (IVRS). [...] The blinded staff were not present at	Low "double-blinded procedures with the randomisation managed and recorded at a central location not at the study sites. […] All subjects, investigators, and study personnel were blinded to the content of the study drug, with the exception of the unblinded study personnel (at most sites a study nurse and a back-up study nurse) who were responsible for the following: Randomising the subject on day 0 through the use of an interactive voice recognition system (IVRS). [...] The blinded staff were not present at	Unclear Detection bias was unclear because it was not clearly stated that study personal was blinded at follow-up time-points.	High Missing data 10-20% and not comparable among study arms (i.e. number of missing and reasons for missing data)	High Missing data 10-20% and not comparable among study arms (i.e. number of missing and reasons for missing data)	Unclear No protocol reported, not trial registry entry reported. However, an entry with good match was identifed: Only IRLS was pre-specified under NCT01382901, registered after study completion, other measures not mentioned, therefore unclear selective reporting

Study	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete continuous outcome data (attrition bias) and support for judgement	Incomplete binary data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
		the time of dosing. The blinded staff obtained all of the clinical measurements pertaining to RLS without knowledge of any other measures obtained: for example, serum ferritin or TSAT%."	the time of dosing. The blinded staff obtained all of the clinical measurements pertaining to RLS without knowledge of any other measures obtained: for example, serum ferritin or TSAT%."				
Cho 2016	Low "a random number sequence generated by the Microsoft Excel program"	Low "Both patients and investigators were blinded to the type of treatments. To maintain the blind, the i.v. bottles and lines were covered with foil by the administering nurse, who played no role in the study beyond administering the solutions."	Low "Both patients and investigators were blinded to the type of treatments. To maintain the blind, the i.v. bottles and lines were covered with foil by the administering nurse, who played no role in the study beyond administering the solutions."	Unclear n.r.	Unclear Missing data ≤10% and unclear if comparable between study arms (i.e. number of missing in each study groups and reasons for missing data were not reported)	Unclear Missing data ≤10% and unclear if comparable between study arms (i.e. number of missing in each study groups and reasons for missing data were not reported)	Unclear No protocol found
Davis 2000	Unclear n.r., "individually assigned to study drug using block randomization by a nurse who was independent from	Low "individually assigned to study drug using block randomization by a nurse who was independent from the study. This nurse kept the study	Low "Investigators and patients were blinded to treatment."	Unclear n.r. "Investigators and patients were blinded to treatment."	High missing data >20% in either study arm	High missing data >20% in either study arm	Unclear no protocol found, pre-specified outcomes in methods-section were reported

Study	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete continuous outcome data (attrition bias) and support for judgement	Incomplete binary data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
	the study. This nurse kept the study code in a locked cabinet until the end of the study."	code in a locked cabinet until the end of the study."					
Earley 2009	Unclear n.r.	Unclear n.r.	Low "The Pharmacy wrapped the solution and all tubing with black opaque plastic coverings to prevent the subjects from seeing the color of the solution. Patients were blindfolded during the brief period for setting up the intravenous line thereby ensuring the treatment blind was maintained. The nurse setting up and administrating the solution was not blinded to the treatment, but was specifically instructed not to discuss treatment condition with anyone. One of the investigators (C.J.E.) dealt with all of the medical issues that arose	Low "The Pharmacy wrapped the solution and all tubing with black opaque plastic coverings to prevent the subjects from seeing the color of the solution. Patients were blindfolded during the brief period for setting up the intravenous line thereby ensuring the treatment blind was maintained. The nurse setting up and administrating the solution was not blinded to the treatment, but was specifically instructed not to discuss treatment	Unclear Unclear number of missing data (unclear number of individuals randomised or analysed)	Unclear Unclear number of missing data (unclear number of individuals randomised or analysed)	High Adverse events were monitored, only side effects and adverse effects reported.

Study	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete continuous outcome data (attrition bias) and support for judgement	Incomplete binary data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
			with treatment and therefore would not have always been blind to the treatment options. However this investigator was not involved in any collection, processing or analysis of data until the blind was broken. All other investigators and study coordinators were blind to treatment."	condition with anyone. One of the investigators (C.J.E.) dealt with all of the medical issues that arose with treatment and therefore would not have always been blind to the treatment options. However this investigator was not involved in any collection, processing or analysis of data until the blind was broken. All other investigators and study coordinators were blind to treatment."			
Grote 2009	Unclear n.r.	Low "Central randomization was performed via a webbased system (IT-Coach, Gothenburg, Sweden) using the minimization method to ensure baseline balance	Low " Specific logistics were implemented to keep the study blinded to both patients and study personnel. Infusions were prepared by the local pharmacy, infusion	Unclear Because primary outcome IRLS was not described to be blinded. "Specific logistics were implemented to keep the study	Low Missing data ≤5%	High missing data >20% in either study arm	Low Registery ISRCTN82469428, all pre-specified outcomes were reported

Study	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete continuous outcome data (attrition bias) and support for judgement	Incomplete binary data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
		for the following variables: date of birth, S-ferritin, IRLS score, and B-hemoglobin. Specific logistics were implemented to keep the study blinded to both patients and study personnel. Infusions were prepared by the local pharmacy, infusion bags and disposables were non-transparent. Infusions and blood chemistry results were supervised by personnel otherwise not involved in the care of the patient."	bags and disposables were non-transparent. Infusions and blood chemistry results were supervised by personnel otherwise not involved in the care of the patient."	blinded to both patients and study personnel. Infusions were prepared by the local pharmacy, infusion bags and disposables were non-transparent. Infusions and blood chemistry results were supervised by personnel otherwise not involved in the care of the patient."			
Lee 2014	Unclear n.r.	Unclear n.r.	High "First, this was not a blinded study, and subjects could be aware of the nature of the medication taken."	Unclear n.r.	High missing data >20% in either study arm	High missing data >20% in either study arm	Unclear no protocol found, pre-specified outcomes in methods-section were reported
Trenkwalder 2016	Low "Randomization was performed based on a pre-defined randomization list,	Low "Randomization was performed based on a pre-defined randomization list,	Low "patient- and assessor-blind (the study nurse who administered the	Low "patient- and assessor-blind (the study nurse who administered the	High missing data >20% in either study arm	High missing data >20% in either study arm	High Two protocols available. Trial registery: one outcome is

Study	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete continuous outcome data (attrition bias) and support for judgement	Incomplete binary data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
	stratified per site, and generated by the Sponsor's Biostatistics department, to which only unblinded study staff had access. Patients were allocated a randomization number in accordance with the randomization schedule generated by the Sponsor's Biostatistics department. This number corresponded to a unique envelope containing the treatment assigned. "	stratified per site, and generated by the Sponsor's Biostatistics department, to which only unblinded study staff had access. Patients were allocated a randomization number in accordance with the randomization schedule generated by the Sponsor's Biostatistics department. This number corresponded to a unique envelope containing the treatment assigned. "	treatment was not blinded)"	treatment was not blinded)"			missing: Time to the need for additional non-FCM RLS treatment due to lack or (time-to-event analysis). Supplemental material: all outcomes reported. In addition, QoL measured, but not reported.
Wang 2009	Low A clinical investigative pharmacist, independent from the study, grouped	Unclear A clinical investigative pharmacist, independent from the study, grouped patients using a randomly generated	Low Double blind and "The clinical investigative pharmacist held the randomization code in a	Low "Clinical investigative pharmacist held the randomization code in a locked cabinet	Low Missing data ≤5%	Low Missing data ≤5%	Unclear No study protocol published; all pre-specified outcome from

Study	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete continuous outcome data (attrition bias) and support for judgement	Incomplete binary data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
	patients using a randomly generated sequenced number program. The clinical investigative pharmacist held the randomization code in a locked cabinet until the end of the study.	sequenced number program. The clinical investigative pharmacist held the randomization code in a locked cabinet until the end of the study.	locked cabinet until the end of the study."	until the end of the study."			method section were reported
Women with fatigue							
FERRIM (Krayenbuehl 2011)	Unclear "The randomization schedule was generated by Cardinal Health Germany GmbH (Schorndorf, Germany)."	Low "The control group received placebo (0.9% saline). It was ensured through organizational measures that neither the patient nor the investigator could become aware of whether the active group" […] "The study medication was prepared and administered by a staff member other than the investigator. Both the infusion bag and the injection site were covered and nontransparent tubing	Low "The control group received placebo (0.9% saline). It was ensured through organizational measures that neither the patient nor the investigator could become aware of whether the active group" […] "The study medication was prepared and administered by a staff member other than the investigator. Both the infusion bag and the injection site were covered and nontransparent tubing	Unclear n.r.	Low Missing data ≤10% and comparable among study arms (i.e. number of missing and reasons for missing data)	Low Missing data ≤10% and comparable among study arms (i.e. number of missing and reasons for missing data)	Low Registered: ISRCTN78430425, all pre-specified patient relevant outcomes reported

Study	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete continuous outcome data (attrition bias) and support for judgement	Incomplete binary data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
		was used, ensuring that the patient could not see the infusion solution at any time. The investigator was not present during the infusion." [...] "The physician's assistant took all necessary precautions to ensure that the patient could not see nor draw any conclusion as to the nature of the solution administered. The work of the physician's assistant was done completely independently of the study physicians."	was used, ensuring that the patient could not see the infusion solution at any time. The investigator was not present during the infusion." [...] "The physician's assistant took all necessary precautions to ensure that the patient could not see nor draw any conclusion as to the nature of the solution administered. The work of the physician's assistant was done completely independently of the study physicians."				
PREFER (Favrat 2014)	Low "computer-generated list of random numbers"	Low "Investigators received a set of sealed envelopes that corresponded to a randomization number and contained the identity of the study drug, and prepared and administered the study drug. Patients were blinded to the study	Unclear Blinding of study personal not guaranteed	Unclear n.r.	Unclear Missing data ≤10% and unclear if comparable between study arms (i.e. number of missing in each study groups and reasons for	Unclear Missing data ≤10% and unclear if comparable between study arms (i.e. number of missing in each study groups and reasons for	Low Protocol provided: all outcomes were reported as they were pre-specified

Study	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete continuous outcome data (attrition bias) and support for judgement	Incomplete binary data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
		treatment by covering infusion bags with opaque bags and using dark-colored infusion lines."			missing data were not reported)	missing data were not reported)	
Vaucher 2012	Low "computer generated"	Low "Each drug package was coded with a unique number according to the randomization schedule and was sent to the relevant practice. General practitioners enrolled the patients and gave them sequentially numbered containers."	Low "The allocation remained concealed to patients, general practitioners, caregivers and principle investigators until the end of the trial. During the analyses, the statistician remained blinded as to what treatment each group received."	Low "The allocation remained concealed to patients, general practitioners, caregivers and principle investigators until the end of the trial. During the analyses, the statistician remained blinded as to what treatment each group received."	Low Missing data 10-20% , comparable among study arms (i.e. number of missing and reasons for missing data) and adequate method used to deal with missing data in the analysis (ex. Multiple Imputation, but not “last observation carried forward”)	Unclear Missing data 10-20% and unclear if adequate methods were used to deal with missing data in the analysis	Unclear No protocol found, all pre-specified outcomes were reported
Verdon 2003	Unclear n.r.	Low "Patients, caregivers, and investigators were blinded to treatment	Low "Patients, caregivers, and investigators were blinded to treatment	Low "Patients, caregivers, and investigators were blinded to	Low Missing data ≤10% and comparable	Low Missing data ≤10% and comparable	Unclear No protocol found, side effect were pre-

Study	Random sequence generation (selection bias) and support for judgement	Allocation concealment (selection bias) and support for judgement	Blinding of participants and personnel (performance bias) and support for judgement	Blinding of outcome assessment (detection bias) and support for judgement	Incomplete continuous outcome data (attrition bias) and support for judgement	Incomplete binary data (attrition bias) and support for judgement	Selective reporting (reporting bias) and support for judgement
		assignment until the end of the trial. Each drug package was coded with a unique number according to the randomisation schedule and then posted to the relevant practice. The codes were held by the pharmacist and remained unbroken until the analyses were completed."	assignment until the end of the trial. Each drug package was coded with a unique number according to the randomisation schedule and then posted to the relevant practice. The codes were held by the pharmacist and remained unbroken until the analyses were completed."	treatment assignment until the end of the trial. Each drug package was coded with a unique number according to the randomisation schedule and then posted to the relevant practice. The codes were held by the pharmacist and remained unbroken until the analyses were completed."	among study arms (i.e. number of missing and reasons for missing data)	among study arms (i.e. number of missing and reasons for missing data)	specified but not reported
Children with ADHD							
Konofal 2008	Unclear n.r.	Unclear n.r.	Low "Patients, parents, teachers, and investigators were totally blind to treatment and to biochemical measures during the trial, which allows confidence that the subjective scoring of ADHD symptoms was unbiased."	Low "Patients, parents, teachers, and investigators were totally blind to treatment and to biochemical measures during the trial, which allows confidence that the subjective scoring of ADHD symptoms was unbiased."	High Missing data 10-20% and not comparable among study arms (i.e. number of missing and reasons for missing data)	High Missing data 10-20% and not comparable among study arms (i.e. number of missing and reasons for missing data)	Unclear No protocol found

5.4 Appendix – Supporting information of the individual patient data meta-analysis

Table 34 List of the availability of biomarkers and variables by trials and measured time point

	FERRIM (Krayenbuehl 2011)	**PREFER** (Favrat 2014)	**Vaucher 2012**	**Verdon 2003**
Biomarkers	Time points measured*	Time points measured	Time points measured	Time points measured
Haemoglobin	screening 6 w 12 w	screening 1 w 4 w 8 w	0 w 6 w 12 w	0 w
Haemotocrit	screening 6 w 12 w	screening 1 w 4 w 8 w	0 w 6 w 12 w	0 w
RBC count	n.r.	screening 1 w 4 w 8 w	0 w 6 w 12 w	0 w
Reticulocytes	n.r.	screening 1 w 4 w 8 w	n.r.	n.r.
Mean corpuscular volume	screening 6 w 12 w	screening 1 w 4 w 8 w	0 w 6 w 12 w	0 w
Mean corpuscular haemoglobin	screening 6 w 12 w	screening 1 w 4 w 8 w	n.r.	n.r.
Mean corpuscular haemoglobin concentration	screening 6 w 12 w	n.r.	n.r.	n.r.
Serum iron	0 w 6 w 12 w	n.r.	0 w 6 w 12 w	n.r.
Soluble transferrin receptor	n.r.	screening 1 w 4 w 8 w	0 w 6 w 12 w	n.r.
Transferrin	0 w 6 w 12 w	n.r.	0 w 6 w 12 w	n.r.
Transferrin saturation	0 w 6 w 12 w	screening 1 w 4 w 8 w	n.r.	n.r.
Serum ferritin	0 w 6 w 12 w	screening 1 w 4 w 8 w	0 w 6 w 12 w	0 w 4 w

C-reactive protein	0 w 6 w 12 w	screening	0 w	n.r.
Total iron binding capacity	n.r.	n.r.	0 w 6 w 12 w	n.r.
Thyroid stimulating hormone	n.r.	n.r.	0 w	n.r.
Further variable				
Fatigue severity	Brief Fatigue Inventory 0 w 6 w 12 w	Piper Fatigue Scale 0 w 1 w 4 w 8 w	Multidimensional Assessment of Fatigue 0 w 12w CAPPS 0 w 12 w	Visual analogue scale 0 w 4 w CAPPS 0 w 4 w
Study center IDs	n.r.	n.r.	centre	n.r.
Age	Only age group by five years	Only age group by five years	n.r.	Age in years
Depression	n.r.	screening	0 w 12 w	0 w 4 w
QoL	n.r.	0 w 8 w	0 w 12 w	n.r.

*in weeks. Abbreviations: n.r., not reported; w, weeks

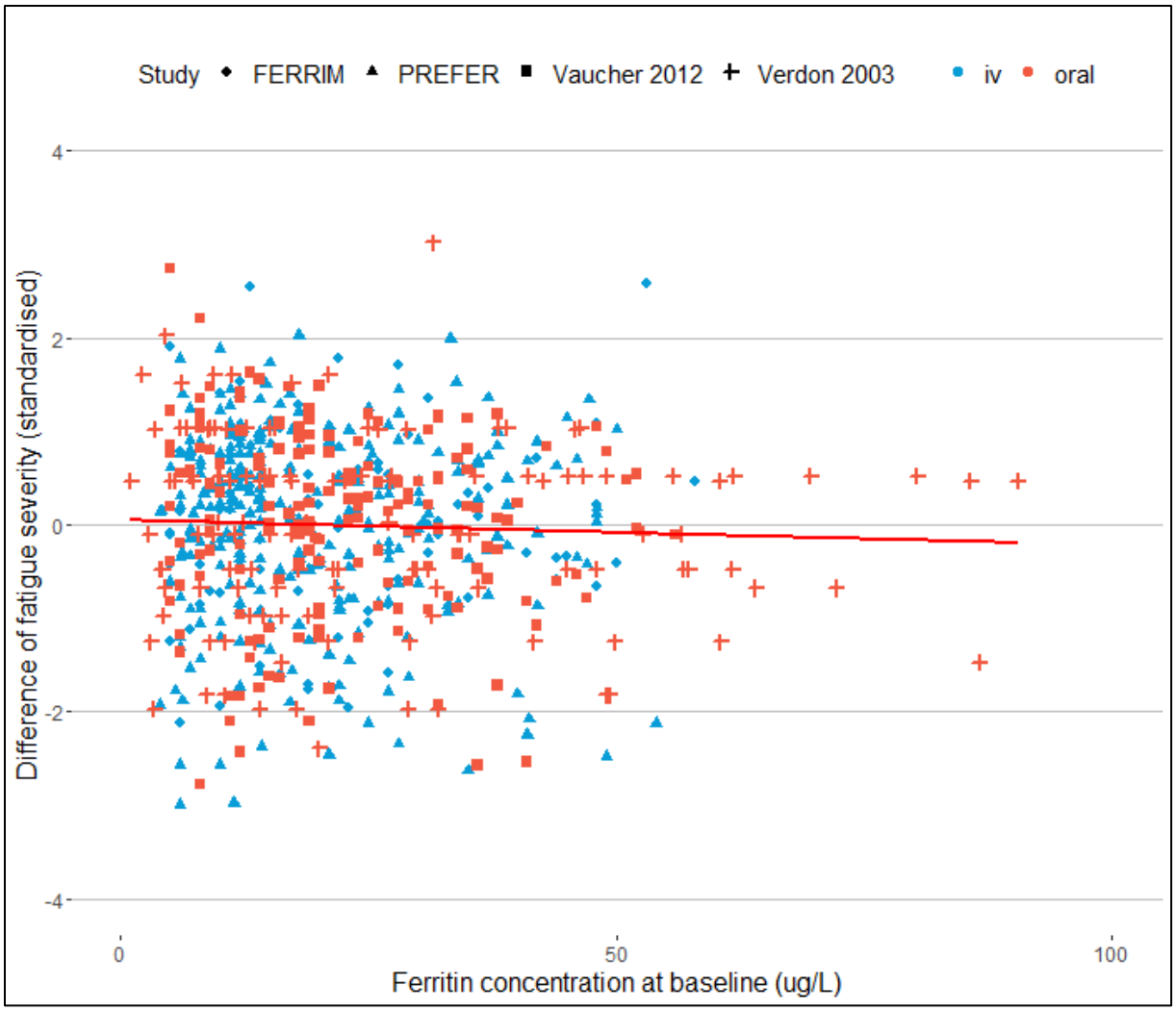


Figure 40 Trial-specific scatterplot with fitted linear unadjusted regression for difference in fatigue severity (standardised) and baseline ferritin concentration as continuous variable excluding ferritin concentrations >100 µg/l

Table 35 Individual patient data meta-analysis – Sensitivity analyses of the multilevel linear regression model for difference in fatigue severity (standardised) and ferritin as continuous variable

Excluding outliers (n=5): Verdon 2003 recruited five women with ferritin concentrations >100 µg/l (see also Figure 40**)**		
Parameter	**Estimates (95% CI)**	**P-value**
Group (intervention vs. control)	0.38 (-0.51 to -0.21)	<0.001
Ferritin concentration at baseline (µg/l)	0.00 (-0.00 to 0.01)	0.358
Follow-up in days	0.00 (-0.00 to 0.00)	0.426
Route of administration (parenteral vs. oral)	0.00 (-0.15 to 0.15)	0.983
Parenteral iron therapy only (n=359)		
Parameter	**Estimates (95% CI)**	**P-value**
Group (intervention vs. control)	-0.33 (-0.54 to -0.13)	0.001
Ferritin concentration at baseline (µg/l)	0.00 (-0.01 to 0.01)	0.358
Follow-up in days	-0.03 (-0.04 to -0.04)	0.007
Oral iron therapy only (n=298)		
Parameter	**Estimates (95% CI)**	**P-value**
Group (intervention vs. control)	-0.39 (-0.61 to -0.17)	<0.001

Ferritin concentration at baseline (µg/l)	0.00 (-0.00 to 0.01)	0.660
Follow-up in days	0.00 (-0.00 to 0.00)	0.617
FERRIM (Krayenbuel 2011) association assessed on fatigue severity original scale Brief Fatigue Inventory (n=75)		
Parameter	**Estimates (95% CI)**	**P-value**
Group (intervention vs. control)	-0.18 (-0.95 to 0.60)	0.658
Ferritin concentration at baseline (µg/l)	0.01 (-0.02 to 0.04)	0.365
Follow-up in days	-0.05 (-0.10 to 0.00)	0.060
PREFER (Favrat 2014) association assessed on fatigue severity original scale Global fatigue index - MAF (n=284)		
Parameter	**Estimates (95% CI)**	**P-value**
Group (intervention vs. control)	-0.85 (-1.33 to -0.37)	<0.001
Ferritin concentration at baseline (µg/l)	-0.01 (-0.03 to 0.01)	0.445
Follow-up in days	-0.10 (-0.17 to -0.02)	0.008
Vaucher 2012 association assessed on fatigue severity original scale 22-item Piper Fatigue Scale (n=163)		
Parameter	**Estimates (95% CI)**	**P-value**
Group (intervention vs. control)	-4.36 (-7.76 to -0.96)	0.012
Ferritin concentration at baseline (µg/l)	-0.06 (-0.20 to 0.09)	0.425
Follow-up in days	-0.03 (-0.15 to 0.09)	0.590
Verdon 2003 association assessed on fatigue severity original scale Visual Analogue Scale (n=135)		
Parameter	**Estimates (95% CI)**	**P-value**
Group (intervention vs. control)	-0.72 (-1.35 to -0.10)	0.023
Ferritin concentration at baseline (µg/l)	0.01 (-0.00 to 0.02)	0.187
Follow-up in days	-0.04 (-0.09 to 0.01)	0.083

Table 36 Individual patient data meta-analysis – Associations of further biomarker at baseline and standardised differences of fatigue severity.

Haemoglobin concentration		
Parameter	**Estimates (95% CI)**	**P-value**
Group (intervention vs. control)	-0.36 (-0.51 to -0.21)	<0.001
Haemoglobin concentration at baseline (g/l)	-0.00 (-0.01 to 0.00)	0.825
Follow-up in days	-0.00 (-0.01 to 0.00)	0.327
Route of administration (parenteral vs. oral)	-0.00 (-0.16 to 0.15)	0.961
Haematocrit		
Parameter	**Estimates (95% CI)**	**P-value**
Group (intervention vs. control)	-0.36 (-0.51 to -0.21)	<0.001
Haematocrit at baseline (vol %)	0.00 (-0.03 to 0.03)	0.819
Follow-up in days	0.00 (-0.00 to 0.00)	0.341
Route of administration (parenteral vs. oral)	-0.01 (-0.17 to 0.16)	0942
Mean corpuscular volume		
Parameter	**Estimates (95% CI)**	**P-value**
Group (intervention vs. control)	-0.35 (-0.50 to -0.20)	<0.001
Mean corpuscular volume at baseline (mu^3)	0.01 (-0.01 to 0.02)	0.356
Follow-up in days	-0.00 (-0.01 to 0.00)	0.263
Route of administration (parenteral vs. oral)	0.03 (-0.13 to 0.18)	0.756

5.5 Appendix – Identification of branch probabilities for parenteral iron therapy

On April 18, 2018, the Medline and the Cochrane library were searched regarding branch probabilities of the *parenteral* branch in the decision tree. The initial number of hits was rather low when including only studies of IDNA populations and excluding IDA. Therefore, the literature review of parenteral iron therapy needed to be structured into more than one search strategy. These strategies are described in this section. The first strategy specifically targeted the probability of phlebitis, and the second strategy was focused on lethal HSR.

The *first review strategy* was focused on phlebitis and was undertaken, since the studies identified in the section "Clinical effectiveness" of this HTA report (section 2) did not yield any utilizable information on the according branch probability. The database search was performed with rather wide search terms, as only *"iron deficiency"* and *"phlebitis"* were required for a record to be identified, as shown in Table 37. Nevertheless, only six records were found. By an additional hand search, 17 further studies were added. This hand search included the screening of the references of the records identified via database search for additional studies based on their title (and abstract, if the title did not yield sufficient information). This hand search process was carried out analogously for oral iron therapy. Figure 41 illustrates the screening process. The total of 23 records was screened by the criteria listed in Table 38. A broad definition was used, also allowing for anemic patients and retrospective studies of data on treatment without randomization of patients. The reasoning behind these criteria was, that the probability of phlebitis should neither be dependent on anaemia, nor on whether a patient is part of a RCT study or not. Also, comorbidities/procedures such as postpartum anaemia, abdominal hysterectomy, and bariatric/gastric surgery were allowed. By contrast, low and middle income countries, where hygienic conditions and education of medical personnel may on average be lower than in Switzerland, were excluded.

Table 37: First search strategy for branch probabilities and number of hits (parenteral iron therapy)

Phlebitis			
Step	**Search terms**	**Medline**	**Cochrane Library**
1	("iron deficiency") AND "phlebitis"	3	3

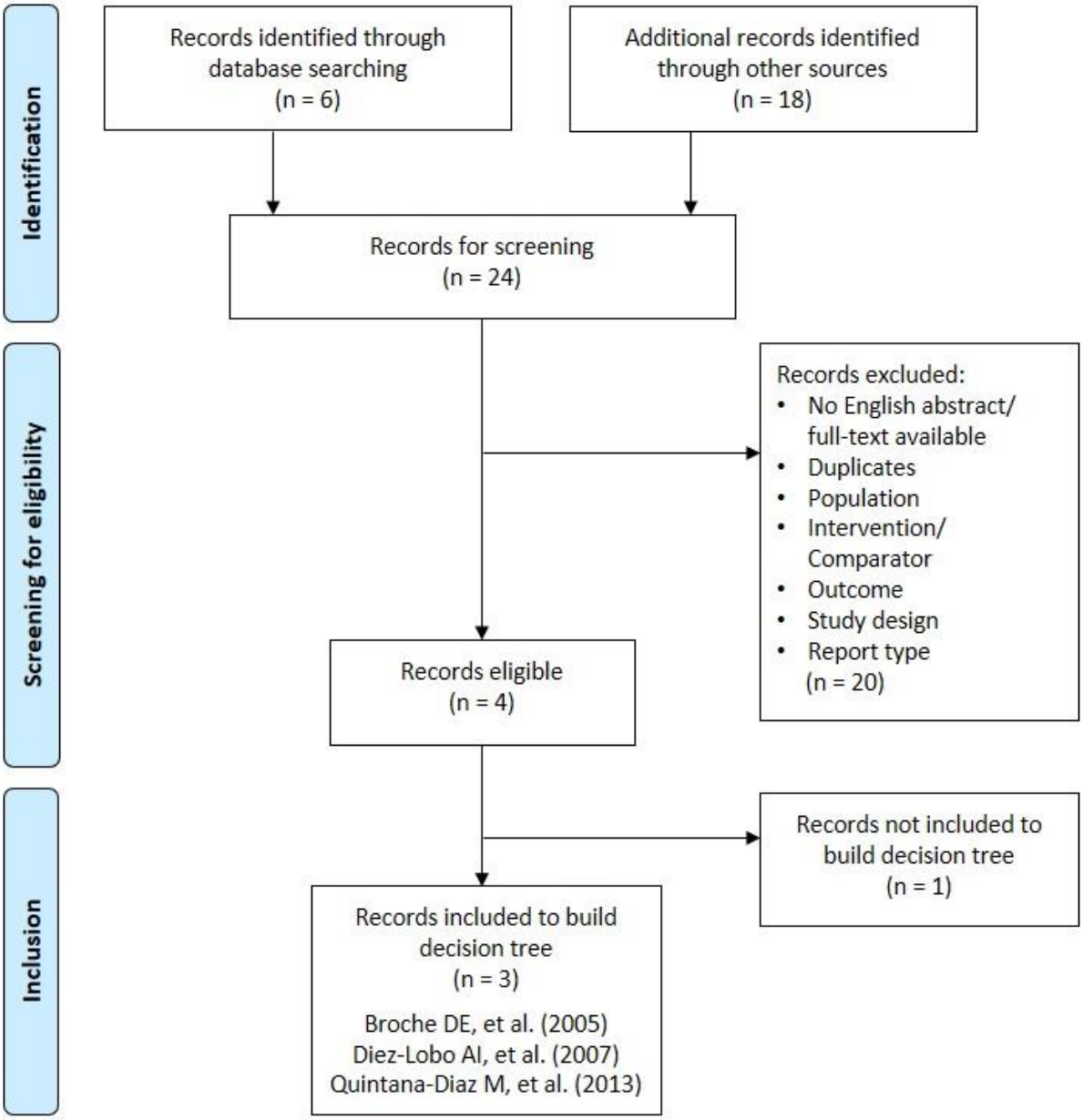


Figure 41: Flow diagram of first search strategy (parenteral iron therapy)

Table 38: Second set of inclusion and exclusion criteria for literature screening (parenteral iron therapy)

Population	Inclusion: • Patients with iron deficiency (IDNA or IDA not necessarily specified) Exclusion: • Low and middle income countries ○ In particular: India, Pakistan
Intervention/ Comparator	Inclusion: • Parenteral iron therapy
Outcome	Inclusion: • Health and safety outcomes
Study design	Inclusion: • Randomized controlled trials (RTC) and quasi-randomized trials • Clinical trials without randomization of patients to multiple groups • Retrospective studies of data on treatment
Report type	Inclusion: • Published articles of study results Exclusion: • Poster presentations and conference abstracts

The *second review strategy* targeted the probability of lethal HSR. The database search terms are shown in Table 39. A total of 176 records resulted from the database search and an additional hand search of the references of the studies identified through the database search. Figure 42 illustrates the screening process. 170 records were excluded, and two studies were used to calibrate the decision tree (base case probability, and upper and lower limit to the probability of a lethal HSR). The inclusion/exclusion criteria are listed in Table 40. These studies are described in further detail in section 3.3.1.

Table 39: Third search strategy for branch probabilities and number of hits (parenteral iron therapy)

Phlebitis			
Step	**Search terms**	**Medline**	**Cochrane Library**
1	("iron deficiency") AND ("hypersensitive" OR "hypersensitivity")	138	32

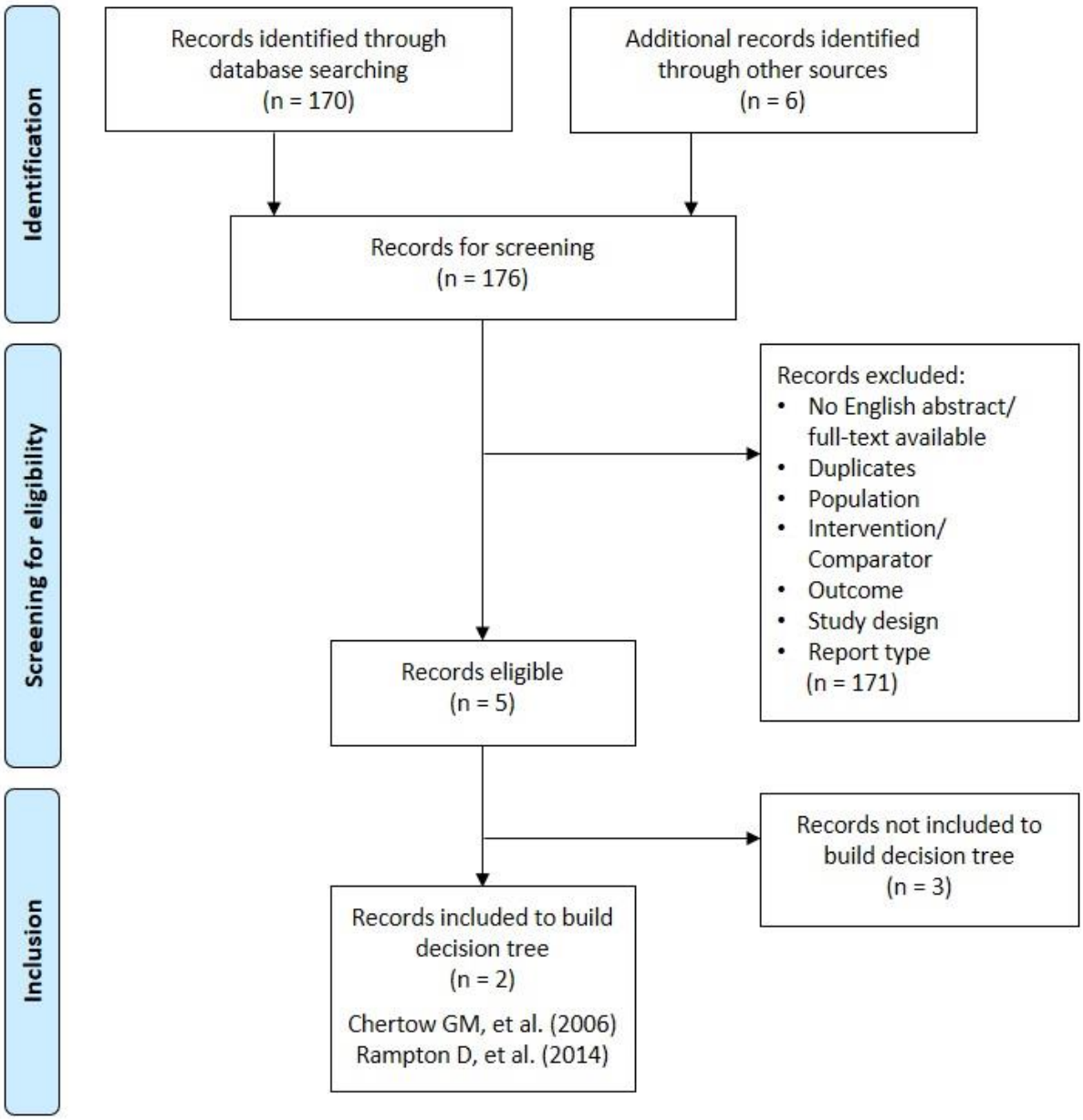


Figure 42: Flow diagram of third search strategy (parenteral iron therapy)

Table 40: Third set of inclusion and exclusion criteria for literature screening (parenteral iron therapy)

Population	Inclusion: • Patients with iron deficiency (IDNA or IDA not necessarily specified) • Lethal HSR for at least one individual Exclusion: • Low and middle income countries
Intervention/ Comparator	Inclusion: • Parenteral iron therapy
Outcome	Inclusion: • Health and safety outcomes
Study design	Inclusion: • Randomized controlled trials (RTC) and quasi-randomized trials • Clinical trials without randomization of patients to multiple groups • Retrospective studies of data on treatment
Report type	Inclusion: • Published articles of study results Exclusion: • Poster presentations and conference abstracts

5.6 Appendix – Identification of branch probabilities for oral iron therapy

The Medline and the Cochrane library were also searched on April 18, 2018, with the aim to find RCTs which provide evidence of the branch probabilities in the decision tree. A moderate number of hits was initially achieved when including only studies of IDNA populations and excluding IDA, and information on the branch probabilities was scarce. Therefore, two different review strategies were applied, regarding both the search terms and the inclusion/exclusion criteria of the screening. These two strategies are described in this section.

The first review strategy was more restrictive. The thereby found branch probabilities were used for the base case estimation of the model. However, the strategy did not yield sufficient information to also construct the lower and upper bounds of the sensitivity analysis. The second review strategy therefore was of more relaxed criteria. It led to the information by which the lower and upper bounds were defined.

The *first review strategy* targeted only patients with IDNA, excluding populations of anaemic patients as well as mixed populations. Table 41 lists the search terms and the number of hits, the latter of which totalled to 56. To supplement the database search, 17 further studies were added based on a hand search. Of these 73 hits, duplicates were removed, and the remaining studies were systematically screened according to the inclusion and exclusion criteria displayed in Table 42. This resulted in 12 studies eligible to serve as a source of branch probabilities of the model (Figure 43 illustrates the inclusion/exclusion process by means of a flow diagram). One out of these 12 studies was used for the calibration of the decision tree, namely for the base case probability of a patient not completing the first cycle of oral treatment and switching to parenteral therapy. One further study supported this calibration by very similar results. The remaining 10 studies met the inclusion and exclusion criteria.

However, they did not provide information which could clearly be interpreted as evidence of the above-mentioned branch probability, or as a probability of any other chance node in the model. For example, labelling of results such as *"number of participants lost to follow up"* did not allow for a precise interpretation in this regard. Section 3.3.1 provides more information on which particular studies were used.

Table 41: First search strategy for branch probabilities and number of hits (oral iron therapy)

Oral therapy			
Step	**Search terms**	**Medline**	**Cochrane Library**
1	("non anemic iron deficiency") AND oral	4	1
2	("non/anemic iron deficiency") AND oral	-	1
3	("nonanemic iron deficiency") AND oral	-	-
4	("latent iron deficiency") AND oral	10	4
5	("iron deficient erythropoiesis") AND oral	23	7
Adverse events			
6	("iron deficient erythropoiesis" OR "latent iron deficiency" OR "non anemic iron deficiency" OR "non/anemic iron deficiency" OR "nonanemic iron deficiency") AND oral AND "adverse event"	-	2
Side effects			
7	("iron deficient erythropoiesis" OR "latent iron deficiency" OR "non anemic iron deficiency" OR "non/anemic iron deficiency" OR "nonanemic iron deficiency") AND oral AND "side effect"	-	1
Malabsorbtion			
8	("iron deficient erythropoiesis" OR "latent iron deficiency" OR "non anemic iron deficiency" OR "non/anemic iron deficiency" OR "nonanemic iron deficiency") AND oral AND "malabsorbition"	-	-
Compliance			
9	("iron deficient erythropoiesis" OR "latent iron deficiency" OR "non anemic iron deficiency" OR "non/anemic iron deficiency" OR "nonanemic iron deficiency") AND oral AND "compliance"	-	1
Adherence			
10	("iron deficient erythropoiesis" OR "latent iron deficiency" OR "non anemic iron deficiency" OR "non/anemic iron deficiency" OR "nonanemic iron deficiency") AND oral AND "adherence"	-	2

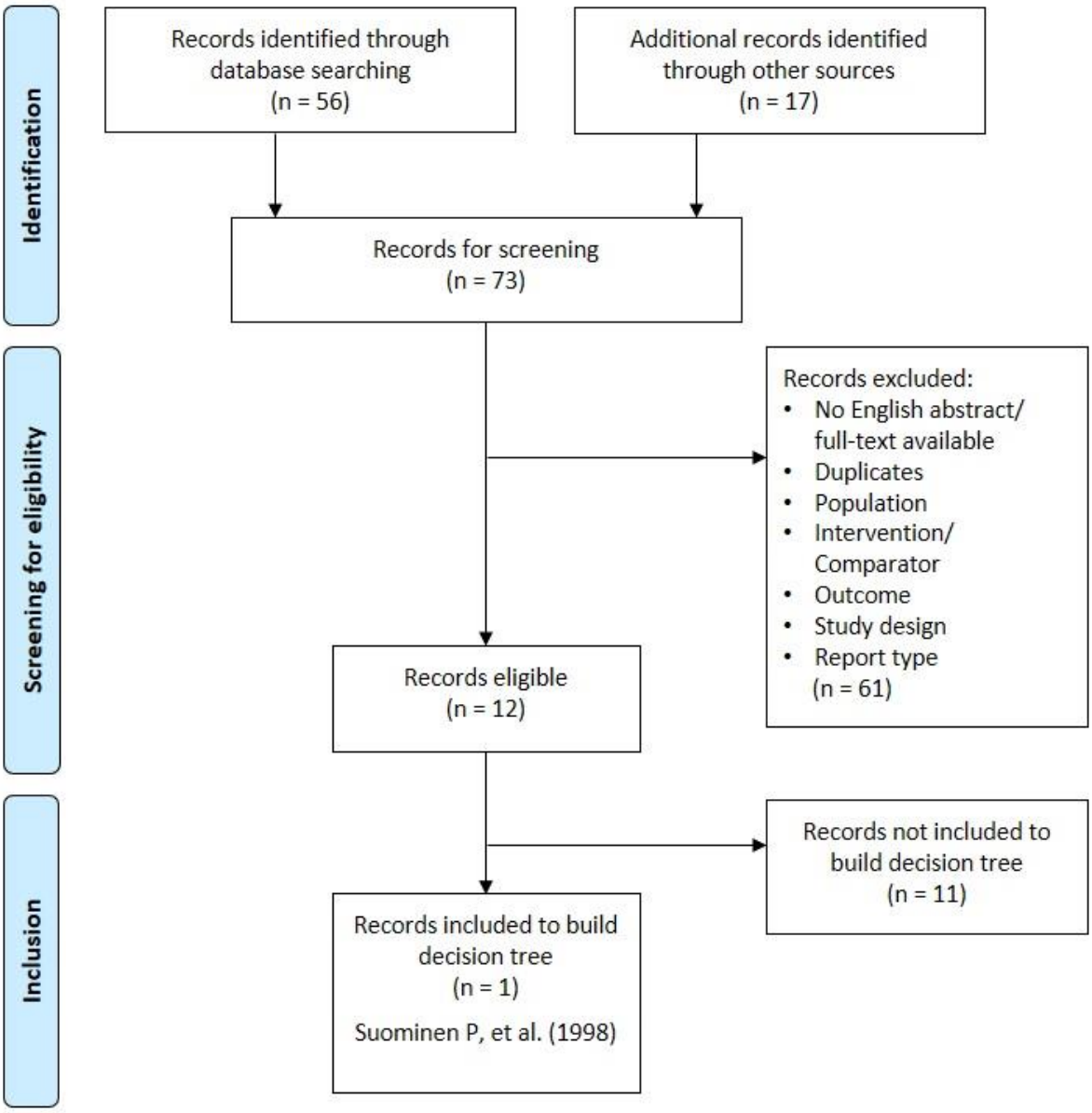


Figure 43: Flow diagram of first search strategy (oral iron therapy)

Table 42: First set of inclusion and exclusion criteria for literature screening (oral iron therapy)

Population	Inclusion:
	• Adults (≥18 years) with IDNA
	Exclusion:
	• Elderlies
	• Low and middle income countries
	  ○ In particular: African countries, Bangladesh, Chile, Mexico, Pakistan
	• Athletes
	• Blood donors
	• Patients with at least one of the following conditions:
	  ○ Iron deficiency anaemia
	  ○ Renal anaemia, microcytic anaemia, Waldenström macroglobulinemia, hemostatic disorder, herediatry hemorrhagic telangiectasia
	  ○ Pregnancy, postpartum hemorrhage, use of intrauterine devices
	  ○ Chronic heart failure
	  ○ Renal failure, chronic kidney disease, dialysis, renal transplant patients
	  ○ Chronic liver failure
	  ○ Chronic inflammatory diseases in particular inflammatory bowel disease, gastrointestinal tract disease, ulcerative colitis
	  ○ Achlorhydria, atrophic gastritis, gastric resection
	  ○ Acute and chronic infections
	  ○ Malignancy
	  ○ Chronic arthritis, rheumatoid arthritis
	  ○ Celiac disease
	  ○ COPD
	  ○ Asymptomatic giardiasis

Intervention/ Comparator	Inclusion: • Oral iron therapy (as intervention or comparator therapy)
Outcome	Inclusion: • Health and safety outcomes
Study design	Inclusion: • Randomized controlled trials (RTC) and quasi-randomized trials • Clinical trials without randomization of patients to multiple groups Exclusion: • Not a primary study • Pilot study
Report type	Inclusion: • Published articles of study results Exclusion: • Poster presentations and conference abstracts

The *second review strategy* allowed for studies of anaemic patients (IDA) and for mixed populations (IDNA and IDA), thereby targeting a wider range of literature than the first step. Table 43 lists the database search terms and the number of hits. The search terms did not specify the form of administration, being oral or parenteral iron therapy, but rather combined the broad term of *"iron deficiency"* with terms referring to side effects and compliance/adherence. Unsurprisingly, this step led to a considerably larger number of hits to be screened. It was undertaken, since the above-mentioned first review strategy revealed the literature to be rather thin regarding the information searched for. 25 further studies were added to the screening process via hand search of the references of the identified studies. Figure 44 illustrates the review process by means of a flow diagram.

Compared to the first set of database search terms shown in Table 41, anaemic populations were included in the second review strategy. As a consequence, several of the identified records stemmed from low and middle income countries or concerned populations with comorbidities, which had not occurred in terms of the first search strategy, and which had to be excluded. Also, the records which were already identified as eligible according to the first review strategy were excluded in order to prevent repetition. Table 44 presents the inclusion/exclusion criteria of the second review strategy. A total of four studies were eligible to serve as sources of branch probabilities of the model. Two of these studies provided the upper and lower bound, respectively, to the probability of a patient not completing the first cycle of oral treatment and switching to parenteral therapy. These studies are discussed in further detail in section 3.3.1.

Table 43: Second search strategy for branch probabilities and number of hits (oral iron therapy)

Iron deficiency			
Step	**Search terms**	**Medline**	**Cochrane Library**
1	("iron deficiency") AND "adverse event"	38	5
2	("iron deficiency") AND "side effect"	53	17
3	("iron deficiency") AND "compliance"	370	2
4	("iron deficiency") AND "adherence"	193	7

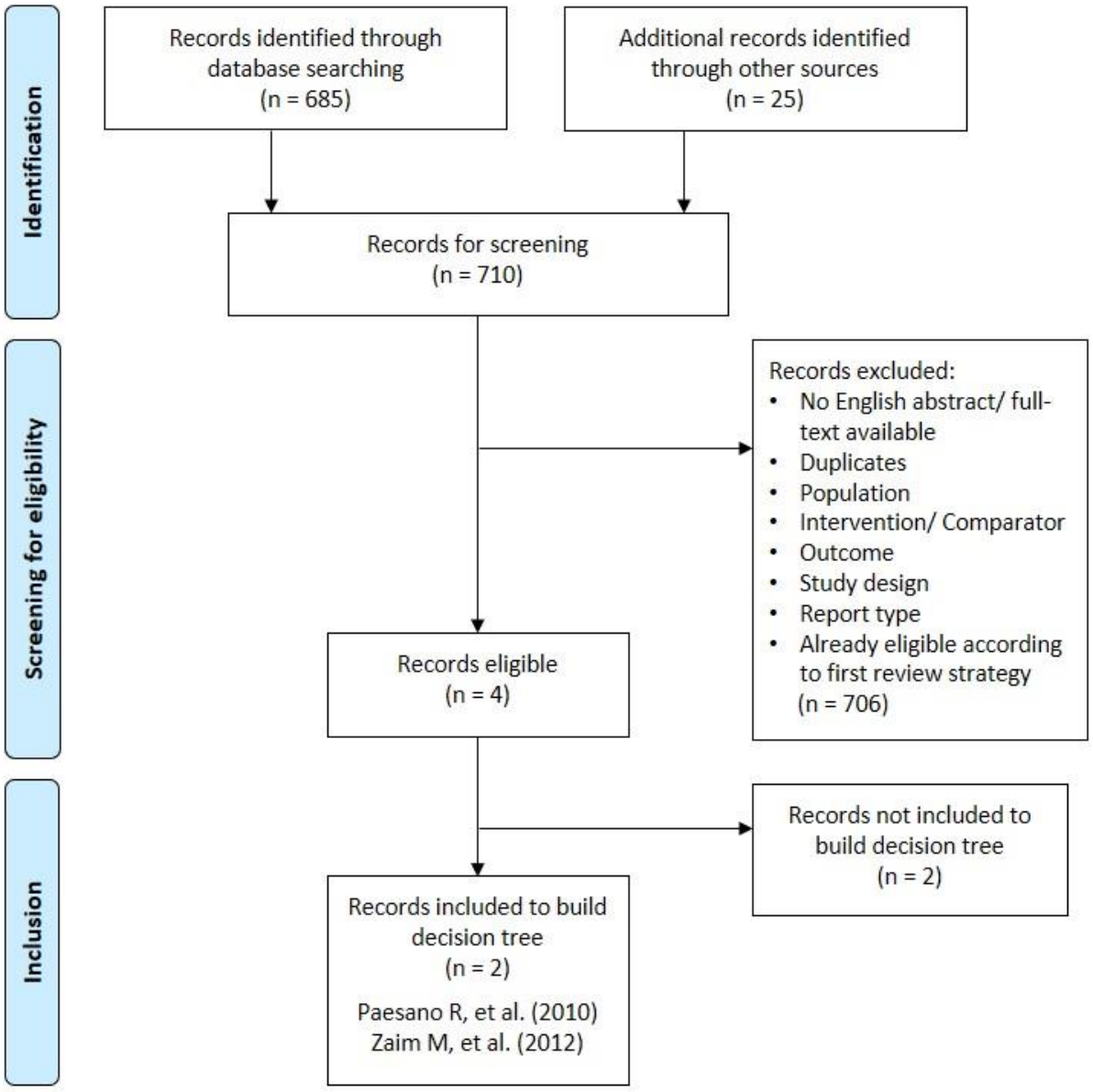


Figure 44: Flow diagram of second search strategy (oral iron therapy)

Table 44: Second set of inclusion and exclusion criteria for literature screening (oral iron therapy)

Population	Inclusion:
	• Adults (≥18 years) with iron deficiency (IDNA or IDA not necessarily specified)
	Exclusion:
	• Elderlies
	• Low and middle income countries
	  ○ In particular: African countries, Bangladesh, Cambodia, Chile, Colombia, India, Mexico, Nepal, Pakistan, Peru, Thailand, Vietnam
	• Athletes
	• Blood donors
	• Homeless people
	• Patients with at least one of the following conditions:
	  ○ Iron deficiency anaemia
	  ○ Renal anaemia, microcytic anaemia, Waldenström macroglobulinemia, hemostatic disorder, herediatry hemorrhagic telangiectasia, hypophosphatemia, sickle cell disease
	  ○ Pregnancy, postpartum hemorrhage, puerperium, use of intrauterine devices, lactating women
	  ○ Chronic heart failure
	  ○ Renal failure, chronic kidney disease, dialysis, renal transplant patients
	  ○ Chronic liver failure
	  ○ Chronic inflammatory diseases in particular inflammatory bowel disease, gastrointestinal tract disease, ulcerative colitis, gastric bypass surgery, autoimmune gastrics, bariatric surgery
	  ○ Achlorhydria, atrophic gastritis, gastric resection
	  ○ Acute and chronic infections
	  ▪ In particular: Malaria, Hepatiis C, HIV
	  ○ Malignancy
	  ▪ In particular: Gastric cancer, chronic myeloproliferative disorders
	  ○ Chronic arthritis, rheumatoid arthritis
	  ○ Celiac disease
	  ○ COPD
	  ○ Asymptomatic giardiasis
	  ○ Obesity
	  ○ Diabetes
	  ○ Neuroleptic akathisia

Intervention/ Comparator	Inclusion: • Oral iron therapy (as intervention or comparator therapy)
Outcome	Inclusion: • Health and safety outcomes
Study design	Inclusion: • Randomized controlled trials (RTC) and quasi-randomized trials • Clinical trials without randomization of patients to multiple groups Exclusion: • Not a primary study • Pilot study

Report type	Inclusion: • Published articles of study results Exclusion: • Poster presentations and conference abstracts

5.7 Appendix – Cost components details

5.7.1 Resource use

Based on input from the **clinical experts**, the following assumptions were made:

- Oral iron therapy consists of a dosage of 100 mg per day for 90 days per treatment cycle. This dosage is lower than recommended by Martius (2009), who suggests 80-100 mg per day for the first week and 200 mg per day for the rest of the cycle[6]. However, the clinical experts made the experience that hardly any patients tolerate 200 mg per day. The prescribing information differs in their recommendation between oral iron drugs included in the specialty list issued by the SFOPH (Table 43). However, the 100 mg per day recommended by the clinical experts seem to be a good approximation of the average of the different recommendations.

Table 45 Dosage recommendation from prescribing information per oral iron drug

Drug	Iron per tablet/capsule	Recommendation
Duofer®	69 mg	1-2 tablets per day
Ferro sanol®	100 mg	1-2 capsules per day
Ferrum Hausmann®	100 mg	Normally 1 capsule per day, in case of severe iron deficiency 2-3 capsules per day
Kendural®	105 mg	1 tablet per day
Maltofer®	100 mg	>12 years old and IDNA: 50-100 mg per day
Tardyferon®	80 mg	1 capsule per day

- Parenteral iron therapy consists of a dosage of maximal 500 mg per infusion. Based on the input from the clinical experts it was further assumed that 40% of the patients have two infusions per cycle and receive the second infusion 1-3 weeks after the first infusion. Consequently, 60% of the patients receive 500 mg and 40% 1000 mg per cycle. The average was 700 mg per patient. This dosage is in line with the recommendation by Martius (2009), who suggests 500-1000 mg[6]. It is also in line with the recommendation by Fehr et al. (2009), who suggest 1000 mg for ferritin concentrations <10 µg/l, and 500 mg for ferritin concentrations between 10-30 µg/l[4]. The prescribing information states that the cumulative iron dosage should be calculated according to the Ganzoni formula: total iron deficit [mg] = cumulative iron dosage [mg] = body weight [kg] x (target Hb – actual Hb) [g/dl] * 2.4 + iron depot [mg]. In case of IDNA the actual Hb is equal to the target Hb. Therefore, the dosage solely depends on the iron depot. The prescribing information recommends an iron depot of 500 mg for a body weight ≥35 kg. The recommendation for Ferinject® is a maximum of 1000 mg iron or 20 mg iron per kg body weight per day. For Venofer®, the maximum per infusion is 500 mg. The 40% of patients who receive 1000 mg per cycle (2 infusions of 500 mg) receive more than recommended according to the Ganzoni formula. However, this may be justified according to the clinical experts as these patients suffer from a chronic imbalance of iron metabolism.

- An office visit is required for each parenteral iron administration. According to the clinical experts, the GP sees the patient for 10 minutes and the patient is monitored by a nurse for a total of 30 minutes during and after the infusion. The prescribing information includes recommendations for the infusion time. For Ferinject®, 200 to 500 mg iron can be injected with a rate of 100 mg iron per minute. Dosages between 500 mg and 1000 mg should be applied over a time of 15 minutes. The infusion time for Venofer® is longer and has been summarized in Table 46.

Table 46 Venofer® infusion time according to prescribing information

Venofer® dosage	Minimal infusion time
100 mg	15 min
200 mg	30 min
300 mg	1.5 h
400 mg	2.5 h
500 mg	3.5 h

  As Ferinject® is much more used than Venofer® (86.3% vs. 13.7%[122]) the 30 minutes monitoring time suggested by the clinical experts seems to cover the average infusion time according to the recommendations from the prescribing information (average infusion time for 500 mg iron: 33 minutes (5 min * 0.863 + 210 min * 0.137)).

- The following material is required per infusion: one IV line, one needle, one syringe and one NaCl 0.9% rinsing solution.

- The follow-up visit lasts 15 minutes. 20% of the patients in either treatment strategy do not return to the GP for a follow-up visit during the first treatment cycle, and hence were not eligible for second cycle.

- The ferritin concentrations is measured during follow-up visit in 80% of the patients. This seems to be in line with the results from Biétry et al. (2017)[98]. The opinions of the clinical experts differ regarding the hemogram ("kleines Blutbild"). Whereas three clinical experts routinely perform one hemogram, one experts does not. In the base case scenario, it was therefore assumed that a hemogram is performed at the follow-up visit. As this is not in line with the results from Biétry et al. (2017)[98] this aspect was further addressed in the univariate sensitivity analysis (section 3.3.4).

- Adverse events for parenteral iron therapy during administration:
  - Mild/moderate HSR: Patients require additional supervision by the GP for 5 minutes and a prolonged infusion time (45 minutes of monitoring by nurse in total).
  - Severe HSR: Leads to inpatient treatment with ICD-10 T88.6 (anaphylactic shock due to undesirable side effect after medication)
  - Phlebitis: Treatment with pain and anti-inflammatory drugs (1 package of Ibuprofen and 50g Venugel) plus one additional office visit with 15 minutes duration.
  - Lethal HSR: No information about the costs of lethal HSR was found in the literature and therefore it was assumed that lethal HSR is associated with inpatient treatment of an anaphylactic shock due to undesirable side effect after medication (ICD-10 T88.6)

5.7.2 Drug costs: oral therapy

Drug	Biggest package size	Package price	Lowest price/mg
Duofer®	100 pc/69 mg	CHF 27.60	CHF 0.00400
Ferro sanol®	50 pc/100 mg	CHF 20.15	CHF 0.00403
Ferrum Hausmann®	100 pc/100 mg	CHF 31.50	CHF 0.00315
Kendural®	90 pc/105 mg	CHF 22.35	CHF 0.00237
Maltofer®	100 pc/100 mg	CHF 35.90	CHF 0.00359
Tardyferon®	100 pc/80 mg	CHF 25.95	CHF 0.00324
Mean			**CHF 0.00340**
Median			**CHF 0.00342**

5.7.3 Drug costs: parenteral therapy

Drug	Package size	Package price	Lowest price/mg	Market share
Venofer®	100mg/5ml/5 amp	CHF 137.95	CHF 0.27590	13.7%
Ferinject®	500mg/10ml/5 amp	CHF 821.45	CHF 0.32858	86.3%
Weighted mean			**CHF 0.32136**	**100%**

5.7.4 GP visit follow-up and lab

Tarmed Position	Description	AL (in TP)	TL (in TP)	CHF using national weighted average TPW	Sources
Costs for GP follow-up visit					
0.0010	Konsultation, erste 5 Min. (Grundkonsultation)	10.42	8.19	16.46	TP: Tarmed Catalogue Tarif 001 - TARMED 1.09, 1.1.2018 TPW: NewIndex, Werte per 1.1.2018
0.0020	+ Konsultation bei Personen über 6 Jahren und unter 75 Jahren, jede weiteren 5 Min. (Konsultationszuschlag)	10.42	8.19	16.46	
0.0030	+ Konsultation, letzte 5 Min. (Konsultationszuschlag)	5.21	4.1	8.23	
Total costs for GP in CHF:					
			10 min visit	24.69	
			15 min visit	41.15	base case
			20 min visit	57.60	
Lab costs					
0.0715	Punktion, venös, zwecks Blutentnahme, jede Lokalisation durch nichtärztliches Personal	0	8.19	7.24	TP: Tarmed Catalogue Tarif 001 - TARMED 1.09, 1.1.2018 TPW: NewIndex, Werte per 1.1.2018
1370.00	Hämatogramm I mittels automatisierter Methode: Erythrozyten, Leukozyten,			8.00	AL, 1.1.2018

	Hämoglobin, Hämatokrit und Indices				
1314.00	Ferritin			7.90	AL, 1.1.2018
4700.00	Auftragstaxe für Auftragnehmer von externen Aufträgen, pro Auftrag und pro Tag; nur anwendbar durch Laboratorien nach Artikel 54 Absatz 3 KVV			24.00	AL, 1.1.2018
Total lab costs in CHF:					
			Hämatogramm	15.24	Base case: 20%
			Ferritin	39.14	
			Hämatogramm + Ferritin	47.14	Base case: 80%

AL = ärztliche Leistung; TL = technische Leistung; TP = Taxpunkt; TPW = Taxpunktwert

5.7.5 GP visit for iron infusion

Tarmed Position	**Description**	**AL (in TP)**	**TL (in TP)**	**CHF using national weighted average TPW**	**Sources**
Costs for GP visit for iron infusion					
0.0010	Konsultation, erste 5 Min. (Grundkonsultation)	10.42	8.19	16.46	TP: Tarmed Catalogue Tarif 001 - TARMED 1.09, 1.1.2018 TPW: NewIndex, Werte per 1.1.2018
0.0020	+ Konsultation bei Personen über 6 Jahren und unter 75 Jahren, jede weiteren 5 Min. (Konsultationszuschlag)	10.42	8.19	16.46	
0.0030	+ Konsultation, letzte 5 Min. (Konsultationszuschlag)	5.21	4.1	8.23	
0.0855	Gefässzugang, peripherven ös, jeder Zugang, durch nichtärztliches Personal	0	35.29	31.21	
0.137	Nachbetreuung/Betreuung/Überwachung in der Arztpraxis bei Personen über 6 Jahren und unter 75 Jahren, pro 15 Min.	4.17	28.01	28.46	
Total costs for GP and nurse in CHF:					
	10 min Konsultation und 30 min Überwachung			112.81	base case
	10 min Konsultation und 45 min Überwachung			141.27	
	15 min Konsultation und 30 min Überwachung			129.27	
	15 min Konsultation und 45 min Überwachung			157.72	
Costs for material					
MiGeL Positions-Nr	**Bezeichnung**		**Menge**	**HVB**	**Source**
03.04 Material für Infusionstherapie					
03.04.01.00.1	Infusionsschlauch normal		1	4.1	Mittel- und Gegenstände-Liste (MiGeL) vom 1.Januar 2018

03.04.04.00.1	Luer-lock-Spritze		1	0.45	
03.04.05.00.1	Nadel		1	0.45	
99.11 Spüllösungen					
99.11.01.00.1	Spüllösung NaCl 0.9%		1 Liter	6.95	
Total costs for material:					
Schlauch, Spritze, Nadel, Spüllösung				11.95	

AL = ärztliche Leistung; TL = technische Leistung; TP = Taxpunkt; TPW = Taxpunktwert

5.8 Appendix - Detailed information on AE probability generation

The base case value derives from Favrat et al. (2014), and the trials from Krayenbuehl et al. (2011) and Trenkwalder et al. (2017) were used for the lower and upper bound, respectively[46,52,53]. Favrat et al. (2014) reported an RCT of women with IDNA from Austria, Germany, Sweden, and Switzerland[53]. IDNA was identified if 1. ferritin saturation laid below 50 µg/l and transferrin saturation below 20% or if 2. ferritin saturation laid below 15 µg/l. The intervention was a parenteral treatment with 1000 mg of ferric carboxymaltose within a 250 ml saline solution over a minimum of 15 minutes. The "most common TEAEs" (treatment-emergent adverse events) listed by Favrat et al. (2014) (headache, nasopharyngitis, pyrexia, nausea) potentially qualify for a mild/moderate HSR according to the typology by Rampton et al. (2014)[96]. Consequently, it was assumed that all 37 patients (37/145=25.5%) with mild/moderate TEAEs were relevant for the assessment. Further 3 patients (3/145=2.1%) with severe TEAEs (one patient: nausea, headache, heavy legs, arthralgia, myalgia; one patient: hematoma and 2x discoloration at injection sites; one patient: headache) were also considered as mild/moderate HSR according to the assessment. This sums to 27.6% (40/145) of patients with mild/moderate HSR according to the assessment. This value was used in the base case analysis. Krayenbuehl et al. (2011) presented an RCT in non-anaemic Swiss women. The intervention group of this RCT was treated with 800 mg of iron III hydroxide sucrose within 800 ml of saline solution over a maximum of 40 minutes[52]. Only drug-associated adverse events were listed in detail (nausea, chills, headache, dizziness, chest pain, dysaesthesia, dysgeusia) and which also potentially qualify for a mild/moderate HSR according to the typology by Rampton et al. (2014)[96]. Consequently, it was assumed that the 20.9% of the treated patients (9 individuals out of 43) who experienced a drug-associated adverse event were affected by a mild/moderate HSR according to the assessment. This value was used as lower bound in the sensitivity analysis. Krayenbuehl et al. (2011) reported that the placebo group had a significantly smaller rate of drug-associated adverse events (6.4%, 3 individuals out of 47). Trenkwalder et al. (2017) conducted a RCT in women and men with IDNA in Germany. The patients in the intervention group were treated with 1000 mg of ferric carboxymaltose over 15 minutes. The thresholds for the diagnosis of IDNA equalled a ferritin saturation of 75 µg/l (or higher ferritin but a transferrin saturation <20%) and a haemoglobin saturation of 115 g/l for women (125 g/l for men). The following TEAEs were reported that potentially qualify for a mild/moderate HSR according to the typology by Rampton et al. (2014)[96]: headache, nausea, arthralgia, back pain, pruritus, feeling cold, abdominal pain upper. There were 18 AEs reported with one of these HSR. Consequently, it was assumed that a maximum of 31.0% (18/58) of the patients were affected by a mild/moderate HSR according to the assessment. This value was used as upper bound in the sensitivity analysis. No probabilities were extracted from the remaining four studies also identified in the section "Clinical effectiveness" of this HTA report (see section 2). Earley et al. (2009), Allen et al. (2011), and Cho et al. (2016) had relatively small samples of 11, 22, and 32 patients, respectively, and Grote et al. (2009) did not provide detailed information on the frequency of adverse events[37,38,41,42].

The probability for experiencing severe HSR during parenteral iron treatment was parametrized from the prescribing information of Ferinject® where it is stated that anaphylactic HSRs can occur "occasionally", i.e. <1/100 and ≥1/1000. Consequently, 0.1% was used as lower and 1% as upper bound, and a base case value of 0.5% was assumed.

As the information regarding the probabilities for experiencing Phlebitis and lethal HSR could not be identified in the RCTs from the section "clinical effectiveness" of this HTA report (see section 2), additional literature searches were conducted. Details are described in the Appendix 5.4.

Regarding the probability of phlebitis, four records were considered eligible, three of which were utilized. Broche et al. (2005) retrospectively analysed clinical data of 217 women with postpartum anaemia in France (haemoglobin saturation <8 g/dl)[114]. A total of 43 out of these women were treated with Venofer®, while the other patients were treated with blood transfusions or with oral iron. The administered dose of parenteral iron was calculated according to the following formula: Total quantity of iron to be replaced in mg = 2.4 x body weight in kg x (target haemoglobin saturation in g/dl – current haemoglobin saturation in g/dl). It was administered as injections of a maximum of 200 mg per 48 hours. One out of the 43 patients treated with parenteral iron experienced phlebitis. This probability of 2.3% was used to calibrate the base case of the decision tree. The results provided by Diez-Lobo et al. (2007) were utilized to define the upper limit of the probability of phlebitis for the sensitivity analysis, which lay at 6.5% (2 individuals out of 31)[116]. The authors retrospectively assessed data on iron deficient women in Spain receiving parenteral iron before an abdominal hysterectomy. Inclusion criteria were serum ferritin saturation <30 ng/ml, serum iron saturation <50 µg/dl, or a transferrin saturation index <20%. The total preoperative dose of parenteral iron sucrose was calculated as follows: Total quantity of iron to be replaced in mg = 2.4 x body weight in kg x (target haemoglobin saturation in g/dl – current haemoglobin saturation in g/dl) + 500. The target haemoglobin saturation was 14 g/dl. Iron sucrose was administered at doses of 200 mg in 200 ml saline solution every 48 to 72 hours, with a maximum of 600 mg per week for 2-4 weeks. Quintana-Diaz et al. (2017) evaluated data of patients at risk of requiring blood transfusion due to iron deficiency in Spain, with haemoglobin saturation <9 g/dl, but who did not require immediate hospitalization[115]. The total iron dose per patient was calculated by the formula according to Evstatiev et al. (2011) as depicted in Table 47[123]. Ferric carboxymaltose was administered at doses of 500-1'000 mg in 100-200 ml of saline solution over 15 minutes. Out of the 238 patients who were treated and attended the follow-up, 170 of which were women and 68 were men, one experienced a case of phlebitis. This rate of 0.4% was used as the lower bound to the branch probability of phlebitis in the model. Malone et al. (2013) provided a comparative review of five randomized controlled trials regarding the safety of parenteral ferric carboxymaltose[124]. Parenteral iron was administered to patients after bariatric and gastric surgery, with a typical dose across the studies being 15 mg/kg up to a maximum of 750 mg per week, and the highest total per patient being 2'250 mg. However, reporting did not go as far as the frequency of phlebitis, with the exception of one additional study being mentioned in the discussion section. The latter study however concerned a single dose of 2'000 mg of iron dextran and a rather small sample of 23 patients and was therefore not considered suitable to serve as a source for the present analysis.

Table 47 Total dose of parenteral iron according to Evstatiev et al. (2011)

Hb (g/dL)	Body weight <70 kg	Body weight ≥70 kg
≥10	1000 mg	1500 mg
7–10	1500 mg	2000 mg

Five studies were considered in determining the probability of lethal HSR. The base case calibration was defined according to the results published by Rampton et al. (2014)[96]. The authors find an average of about one causal death per 5 million doses of parenteral iron sold between the years of 1979 and 2005 based to death certificate data from the US national Center for Health Statistics. They base these results partly on the report by Chertow and Winkelmayer (2010)[125]. This rate of 0.00002% was used to calibrate the branch probability of a lethal HSR in the base case model. The minimum and maximum limit of the sensitivity analysis was defined according to Chertow et al. (2006), who present the number of deaths per dose of parenteral iron sold based on data from the US Food and Drug Administration MedWatch and from IMS Health[117]. The results are reported specifically for different iron products. The lowest number of deaths per dose was reported for Venofer® with 0.000012% (one death per 8.837 million doses sold). The highest rate was observed for the product Dexferrum with 0.000078% (one death per 1.2815 million doses sold). Of the three further studies, which were eligible but not used for the calibration of the model, Wysowski et al. (2010) report a range for the number of deaths per ICD-10 diagnose code Y44.0 in the US between 2002 and 2006[126]. However, according to the definition of this diagnostic code, it is not clear that all of these deaths occurred due to treatment with parenteral iron. Bailie et al. (2010) and Wetmore (2017) do not indicate explicit numbers of deaths caused by parenteral iron[127,128].

The base case probability of completing the first oral cycle had to be parametrized based on a study identified through an additional literature search (see Appendix 5.6 for details) because the studies identified in the section "clinical effectiveness" of this HTA report (see section 2) did not provide information on the frequency of adverse gastrointestinal events or the information provided was not considered specific enough. In an example, Vaucher et al. (2012) performed a RCT in women suffering from IDNA and symptoms of fatigue in France[56]. 102 individuals underwent a 12-week oral treatment with 80 mg of ferrous sulphate per day. 11.8% (12 individuals out of 102) experienced gastrointestinal disorders. However, the authors do not indicate whether these 12 individuals ceased oral therapy due to the adverse events. In another example, Verdon et al. (2003) undertook an RCT of women with IDNA and symptomatic fatigue in Switzerland[9]. 75 individuals underwent a 4-week oral treatment with 80 mg of ferrous sulfate per day. The result showed that 5.3% (4 individuals out of 75) were *"lost to follow-up"*. However, the number of patients who switched to parenteral iron therapy due to adverse events was not reported. From the additional literature search conducted, 12 records were considered eligible (see Appendix 5.6 for details). The study from Suominen et al. (1998) was finally used. The authors assessed 74 healthy adults from Finland, 49 of whom being women, taking 100 mg of ferrous sulphate daily over an intervention period of 12 weeks[118]. All individuals were considered healthy, with the exception of 40% of women having a condition of IDNA with ferritin <22 µg/l and sTfR>2.75 mg/l. None of the men were iron deficient. The patients were not anaemic, had no other relevant comorbidities, and were not pregnant. The exclusion criteria regarding anaemia was a haemoglobin saturation of <117 g/l in women and <128 g/l in men. 12.2% of the healthy adults (9 individuals out of 74) withdrew from the trial prematurely due to adverse gastrointestinal effects. This percentage was used to calibrate the base case branch probability mentioned above. Among women, the share was 12.2% (6 individuals out of 49), and among men it amounted to 12.0% (3 individuals out of 25). Hence, no systematic relationship between the share and the gender was observed. Also, since 40% of the women had IDNA while none of the men did, there was no evidence that iron deficiency affects the probability of dropping out due to adverse gastrointestinal events. As the other studies identified did not provide information on the frequency of adverse gastrointestinal events or the information provided was not considered specific enough, a further literature search was conducted in which

anaemic and mixed (IDA and IDNA) populations were included (see Appendix 5.6 for details) for parametrizing the lower and upper bound of the probability of completing the first oral cycle. From the four eligible studies identified through this additional search, the results from Zaim et al. (2012) and Paesano et al. (2010) were used[119,120]. Zaim et al. (2012) conducted a RCT of 399 women with IDA in Italy. 201 randomly selected individuals underwent a 12-week oral treatment with 105 mg of ferrous sulfate per day. 9.0% (18 individuals out of 201) discontinued the therapy due to adverse events[119]. Zaim et al. (2012) do not specify, how many of these adverse events were treatment-emergent. This group of 201 individuals was considered the control group, since the intervention was the administration of an innovative drug with a lower dosage of iron and a prolonged release. Paesano et al. (2010) presented a RCT of 180 women in Italy, some of which suffering from IDNA and some of which being anaemic. Of the 90 individuals treated with 520 mg of ferrous sulfate per day, 15.5% (14 individuals) withdrew from the study because of side effects. Patterson et al. (2001) present an RCT of a rather small number of individuals (only 22)[129]. Further, the definitions of the type and severity of side effects did not allow for a clear interpretation in the sense of the branch probability searched for. Leonard et al. (2014), report frequencies of specific adverse events, but the share of individuals suffering from at least one adverse event remains unclear, as an individual may have multiple conditions[130].

Regarding the probabilities based on input from the clinical experts, it was proceeded as follows: In a first step, an extensive interview with one out of the four available clinical experts was conducted. This expert was considered to be most familiar with the current practice in the outpatient setting in Switzerland. This expert made his best guess for the base case value. In a second step, his suggested base case value was validated by the other three clinical experts. Two experts agreed on the suggested base case values and one expert was uncertain. However, as the uncertain expert did not provide any alternative values, the values the other three experts agreed on were used. The lower and upper bounds of the probabilities with base case values stemming from the clinical experts were defined by the authors of the study.