Università degli Studi di Firenze

Dipartimento di Statistica "G. Parenti"

Dottorato di Ricerca in Statistica Applicata

Ciclo XXIII – SECS-S/01

Network Meta-analysis and Inconsistency. Application to surgical treatments of gingival recession

Jacopo Buti

Tutor: *Prof. Annibale Biggeri*

Co-Tutor: *Dr. Michela Baccini*

Coordinatore: *Prof. Fabio Corradi*

Logic will get you from A to B.

Imagination will take you everywhere.

Albert Einstein

a Gaia

Acknowledgment

This thesis was processed in the University of Florence, Department of Statistics “G. Parenti”, in collaboration with Department of Periodontology - School of Dentistry.

First of all, I would like to thank my supervisors, Prof. Annibale Biggeri and Dr. Michela Baccini from Department of Statistics, University of Florence, Biostatistics Unit, Institute of research and cancer prevention (ISPO), Florence.

I would like to thank all the people working with me at the Department of Periodontology - School of Dentistry - University of Florence. In particular: Prof. Giovan Paolo Pini Prato for his guidance and support and Dr. Michele Nieri for his valuable suggestions, warm help and encouragement from the beginning till the end.

Moreover, special thanks to Prof. Helen Worthington and Dr. Anne-Marie Glenny working at the Cochrane Oral Health Group, to Dr. Marco Esposito – School of Dentistry - University of Manchester and Prof. Graham Dunn and all the people working at the Health Metodology Research Group – School of Medicine - University of Manchester for their help and the kindness during my visiting period in Manchester.

Glossary

ADM: Acellular Dermal Matrix

BM: Barrier Membrane

CAF: Coronally Advanced Flap

CAL: Clinical Attachment Level

CRC: Complete Root Coverage

CrI: Credibility Interval

CTG: Connective Tissue Graft

DC: Direct Comparison

EMD: Enamel Matrix Derivative

FE: Fixed-Effect

HF-DDS: Human Fibroblast-Derived Dermal Substitute

IC: Indirect Comparison

KT: Keratinized Tissue

KL: Kullback-Leibler distance

LogOR: Log Odds Ratio

MTC: Mixed Treatment Comparison

NM: Network Meta-analysis

OR: Odds Ratio

PRP: Platelet-Rich Plasma

RecRed: Recession Reduction

RE: Random-Effect

SD: Standard Deviation

SM: Standard Meta-analysis

Contents

Introduction

Meta-analysis is a statistical method for quantitatively synthesizing evidence from multiple trials with the aim of obtaining overall pooled effect estimates. This method was initially developed more than a century ago by Karl Pearson (1904), but in the last 30 years the impact of meta-analysis has grown definitly. Standard meta-analytic methods typically synthesize evidence of studies that compare directly, head-to-head, two (or more) interventions. As the number of available treatments increases, the number of possible pair-wise comparisons increases quadratically. In absence of all possible direct comparisons of treatments from single trials and in order to allow indirect evidence to be included in a single analysis, Network Meta-analysis was developed as an extension to standard meta-analysis. In fact, NM is a technique to combine information from all the randomized comparisons, in the absence of a set of randomized trials directly comparing all possible treatment options, allowing to detect the most effective treatment/intervention. The estimate for the relative effectiveness of two treatments uses both the direct, head-to-head, comparisons and indirect evidence. In this thesis, the statistical methods for carrying out a Network Meta-analysis (NM) are investigated and different Bayesian NM models are specified and compared. Sources of variability are explored and the attention is focused on Inconsistency (Incoherence) which can be viewed as a sort of uncertainty due to discrepancy between direct and indirect inference on pair-wise comparisons of treatments/interventions. An application to periodontal treatments of gingival recession is performed. We specified NM Consistency and Inconsistency models based on the approach proposed by Lu and Ades (2006). We developed a Bayesian extension of the model used by Lumley (2002). Finally the Kullback-Leibler distance is introduced as a way to compare direct inference from standard pair-wise meta-analysis and indirect inference from the NM models.

Problem statement

This thesis aims to construct a Bayesian Network Meta-analysis model in the context of treatments of gingival recession. Different models are compared and the concept of Inconsistency between direct and indirect evidence is investigated.

Disposition

In the chapter 1, the background and the research problem are described in a paper (in press) entitled:

“Network Meta-analysis of Randomized Controlled Trials. Direct and Indirect Treatment Comparison”

In chapter 2, methods used for the statistical analyses are presented.

Chapter 3 describes an application of NM models to a data-set of treatments of gingival recession. Results of the analysis are presented and discussed.

Finally in chapter 4, the conclusions from both a clinical and a statistical point of view are debated and recommendations for future works are given.

The WinBUGS codes are listed in Appendices A-C.

1 Extended Background

1.1 Network Meta-analysis of Randomized Controlled Trials. Direct and Indirect Treatment Comparison.

Buti J, Glenny AM, Worthington H, Nieri M, Baccini M

***Keywords*:** Network meta-anaysis, Indirect comparisons, Mixed treatment comparison.

Introduction

Health care decisions should be based on best available evidence. Meta-analyses of direct, head-to-head, randomized controlled trials (RCTs) generally provides a reliable way to summarize evidence relating to efficacy and safety of comparing treatments and can be considered the gold standard for evaluating the effectiveness of healthcare interventions (Liberati 2009, Nieri 2009).

Meta-analysis is commonly used for summarizing results from a set of different indipendent trials aimed to investigate comparisons of treatments/interventions for a given patient population (Egger 2001). In particular, in the fields of medical research and clinical practice, systematic reviews and meta-analyses are developed to ensure that medical treatments are based on the best available empirical data.

Meta-analysis was developed more than a century ago by Karl Pearson (1904), but in the last 30 years (Glass 1976) the impact of meta-analysis has grown definitly. Researchers in many fields, beginning in the 1980s, have been moving away from the narrative review, adopting systematic reviews and meta-analysis according to the need of evidence based conclusions (Borenstein 2009).

Today, meta-analysis is widely used in clinical trials, epidemiology and evidence based medicine. Main issues of meta-analysis include the capacity of summarizing different findings in an overall estimate enhancing results communication; comparing findings coming from different studies and

investigate heterogeneity sources; increasing the power by increasing the sample size.

Standard meta-analytic methods typically synthesize evidence of studies that compare directly 2 interventions (for example, treatment A versus treatment B) alone. Pair-wise comparisons of two treatments are frequently made in one (or a set of trials) and other comparisons made in other trials, but as the number of available treatments increases, the number of possible pair-wise comparisons increases quadratically.

Thus, in presence of a large number of health-care interventions for the same condition, direct comparisons of specific treatments or regimens of interest may not be available in single randomized controlled trials and lack of all possible comparisons is frequently recognised in the body of literature. The pair-wise comparison may be inapplicable due to it disallowing indirect comparisons, and consequently it is impossible to decide upon the best treatment in a class with no common comparator. For instance, an initial trial compares drug A to drug B, while a different trial, studying a similar patient population, compares drug B to drug C. Head-to-head, comparisons fail to define the relationship between the drug A and the drug C in absence of a trial comparing them directly.

In the absence of a set of randomised trials directly comparing all possible treatment options and in order to choose the most effective treatment if more than two treatments exist for the same disease, we can rely on indirect comparisons of multiple treatments. For example, if we have trials comparing A vs C and trials comparing B vs C an indirect estimate of the benefit of A over B can be obtained (Bucher 1997, Song 2003, Yazdanpanah 2004) even though indirect comparison (Fig.1) produces relatively imprecise estimates (Caldwell 2005).

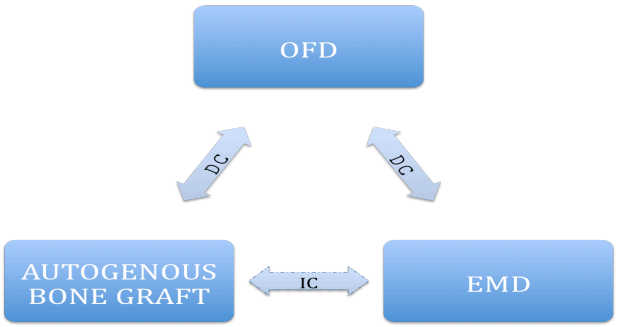


Fig.1 An example of indirect comparison (IC) for the treatment of intrabony defects. Evidence from direct comparison (DC) of autogenous bone graft versus Open Flap Debridment (OFD) in a trial and Enamel Matrix Derivative versus Open Flap Debridment (OFD) in another trial can be combined to have information on the relative treatment effect of autogenous bone graft versus Enamel Matrix Derivative.

Network Meta-analysis (also called Mixed-Treatment Comparisons - MTC - method, Lu 2004) was developed as a new approach to meta-analysis. Differently from standard meta-analytic techniques which allow for single separate pair-wise comparisons, Network Meta-analysis (NM) is able to combine evidence from both direct and indirect comparisons from different trials in a unique network of treatments. Diversity in treatment effects (Salanti 2008) may be present across comparisons in a network, so studies directly comparing treatment A versus treatment C may systematically differ from trials comparing treatment A versus treatment B and, for instance, B versus C from which an indirect estimate of treatment A versus treatment C is obtained. Network Meta-analysis approach is able to consider the inconsistency in the evidence structure, which can be viewed as a sort of uncertainty due to discrepancy between direct and indirect inference on pair-wise comparisons. In the absence of considerable inconsistency, treatment effects can be estimated in a reliable way by taking into account all evidence.

Applications of NM have been recently published in medical journals (Glenny 2005, Psaty 2003, Welton 2009, Orme 2010, Bridle 2003, Wilby 2005). Thus far, in the area of dental sciences, the literature search revealed two applications of Network Meta-analysis (Walsh 2010, Tu 2010).

The aim of the present study was to review current approaches to Network Meta-analysis and discuss different statistical methods to deal with indirect comparisons of treatments and combining direct and indirect evidence. Advantage and disadvantage of these methods are also discussed. Statystical models and specific methods developed for NM are not described in this paper. For further details we remand to Lu and Ades (2004, 2006), Lumley (2002), Higgins (1996), Song (2003), Salanti (2008).

Indirect comparison

Thirty-one (31) of 327 (approximately 9.5%) meta-analyses of RCTs identified through DARE (a recent review of the Database of Abstracts of Reviews of Effects) included some form of indirect comparison (Glenny 2005).

Methods used for indirect comparison have been studied by several authors (Baker 2002, Bucher 1997, Caldwell 2005, Glenny 2005, Lumley 2002, Newhouse 2000). Reliabilty of these approaches was assessed even if literature research revelead a lack of suitable indexing terms in the electronic databases.

Glenny et al. (2005) conducted a systematic review analyzing a large body of literature concerning statistical methods to deal with indirect comparisons of treatments. Throught the search for such papers authors recognised lack of terminology for indirect comparisons; terms used in the papers include cross-study comparison (Phillips 2003, Tsong 2003), connected comparative experiment (Hirotsu 1999), network meta-analysis (Lumley 2002), mixed treatment comparison (Ades 2003), and virtual comparison (Wang 2002). Hardly any of the papers identified cited any of the others.

The authors (Glenny 2005) reviewed the frequency of use of indirect comparisons in systematic reviews and evaluated the methods used in their analysis and interpretation. They concluded that direct evidence from good-quality RCTs should be used wherever possible. Without this evidence, it may be necessary to look for indirect comparisons from RCTs.

Other authors (Edwards 2009, Song 2009, O'regan 2009, Glenny 2005) recently reviewed methods for indirect comparisons and proposed decisional alghoritms (Gartlehner 2008).

Difficoulties and methodological problems related to improperly use of indirect comparisons are strictly related to the presence of biases and confounders. Song et al. (2003) focused on the necessity of similarity between trials involved in the meta-analysis to avoid discrepancy between the direct and indirect estimates.

The Cochrane Collaboration's guidance to authors states that indirect comparisons are not randomized, but are "observational studies across trials, and may suffer the biases of observational studies, for example confounding." (Higgins 2005)

Unadjusted (naive) indirect comparison

Unadjusted indirect comparisons appear as a simple but uncorrect method to summarize evidence from a set of different studies. Data from single arms, part of different trials, are compared naively as if they were from a single (large) controlled trial. As showed in Fig.2, in a first step an unadjusted indirect comparison would combine the summary statistics for study arms for treatment A (formulating a weighted overall summary measure). The same would be done for treatment B. In a second step, a comparison would then be performed on the basis of these two overall summary measures. Data from the comparative arms are not used (Gartlehner 2008). This approach is methodologically flawed because the randomization is broken (Glenny 2005). In these terms, the exclusion of non-randomized studies or observational surveys can be questionable. Gartlehner et al. (2008) suggested that unadjusted indirect comparisons should always be avoided increasing liability to bias and producing overprecise estimates.

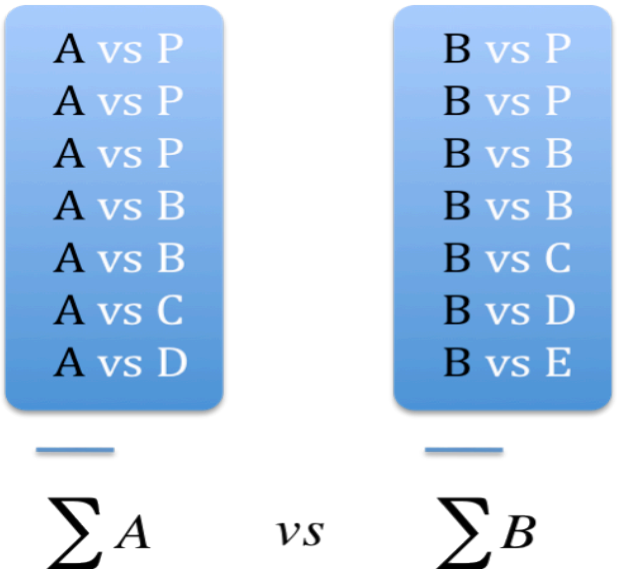


Fig.2 Unadjusted (naive) indirect comparison.

Adjusted indirect comparison

In presence of a common comparator P (placebo, active-control or the gold standard treatment), the relative effects of treatment A versus P (as extracted from one trial) and the relative effects of treatment B versus P (as extracted from another trial) can be compared to obtain an indirect comparison.

Bucher et al. (1997) first described adjusted indirect comparison. For trials with a binary outcome they suggested combining odds ratios from separate meta-analyses, OR_{AB} and OR_{AC}, so that $logOR_{BC}$ is estimated as $logOR_{AB} - logOR_{AC}$, and its variance as $var\ (log\ OR_{BC}) = var\ (logOR_{AB}) + var\ (logOR_{AC})$. From these calculations, it is simple to obtain a confidence interval for $logOR_{BC}$ and hence, by transformation, an estimate of OR_{BC} with a confidence interval. The adjusted indirect comparison method is quite general, and this formulation is clearly a specific example of the general method. Thus, given two estimated effects d_{AB} and d_{AC} for comparisons of A versus B and A versus C, respectively, the effect for the comparison B versus C is estimated as $d_{BC} = d_{AB} - d_{AC}$, and $var\ (d_{BC}) = var\ (d_{AB}) + var\ (d_{AC})$. A 95% confidence interval for d_{BC} is obtained as $d_{BC} \pm 1.96\sqrt{[var(d_{BC})]}$. The method is based on a two-step analysis in which step 1 consists in carrying out single separate meta-analyses and step 2 in combining the first step results in an "adjusted indirect comparison". Fig.3 shows an example of adjusted indirect comparison.

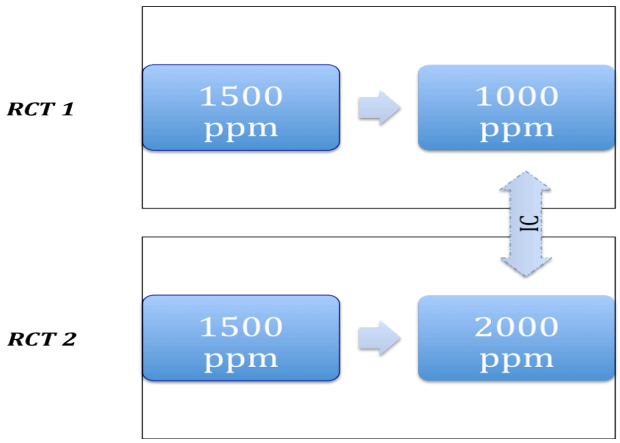


Fig.3 An example of potential indirect comparison (IC) by Walsh et al. (2010). A common comparator (1500 ppm) was used allowing for indirect comparison of different concentrations of fluoride toothpastes of 1000 ppm and 2000 ppm.

Network Meta-analysis

Relative efficacy and effectiveness of healthcare interventions in a network of trials can be evaluated using more complex statistical techniques aimed to combine direct and indirect evidence from different trials in a single analysis.

In 1990, Eddy et al. introduced a Bayesian approach to indirect comparison of treatments (Confidence Profile Method - CPM). The method was the first attempt to combine direct and indirect evidence using data from different sources such as observational studies or trials not accounting for randomization of treatments. High risk of bias was a limitation of this method.

Network Meta-analysis (or Mixed Treatment Comparison) can be considered an evolution of CPM approach, combining information from all the randomized comparisons among a set of several treatments (Gleser 1994, Higgins 1996, Song 2003, Lumley 2002, Psaty 2003, Lu 2004, Lu 2006, Salanti 2008).

Network Meta-analysis allows for:

- Simultaneous comparison of multiple treatments in a single analysis;
- Multiple treatment meta-analysis of randomized controlled trials using direct and indirect evidence;
- Ranking of treatments using a Bayesian approach (Spiegelhalter 2004).

Lumley et al. (2002) proposed a method to carry out an analysis including two-arm trials. Others authors (Ades 2003, Lu 2004, Caldwell 2005) described a framework able to consider in a single analysis any number of treatment groups in multi-arm trials, allowing to make indirect comparisons while fully respecting the randomized structure of the evidence.

In Fig.4 we show a network diagram for the Network Meta-analysis of periodontal treatments of gingival recessions (Cairo 2008) for the RecRed (Recession Reduction) outcome variable. Each node in the network represents a surgical treatment arm from 22 different studies and the solid blue lines between nodes represent pair-wise comparisons for which direct evidence is available from clinical trials. The Network Meta-analysis model was used to evaluate all pair-wise comparisons in the network, including indirect comparisons (dotted yellow lines between nodes) for which no head-to-head data are available from clinical trials.

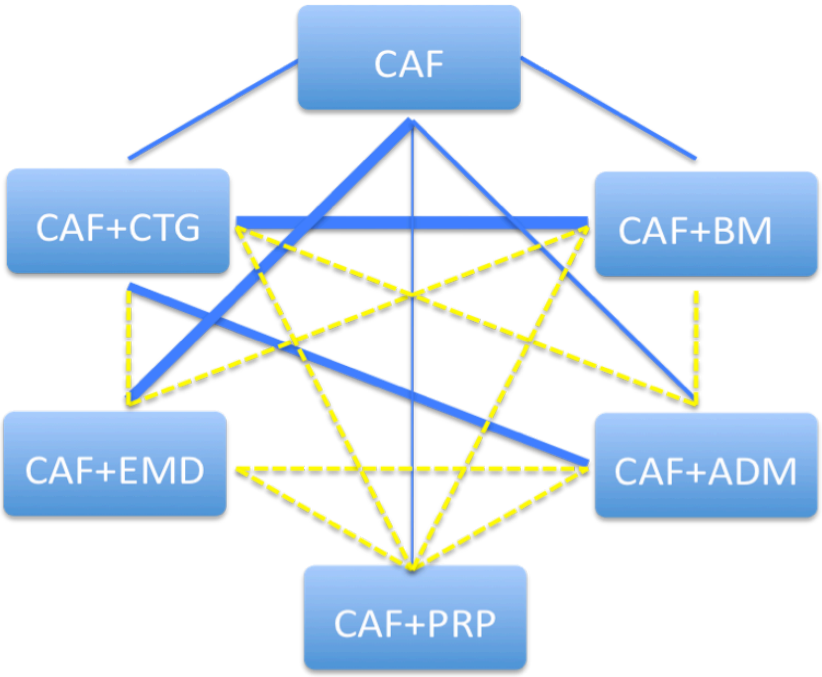


Fig.4 Network diagram for the comparison among 6 different treatments (CAF, CAF+CTG, CAF+BM, CAF+EMD, CAF+ADM, CAF+PRP) of gingival recessions for the RecRed (Recession Reduction) outcome variable. The solid blue lines represent comparison informed by direct evidence from studies. Dotted yellow lines refer to those comparisons that have not been tested directly in randomized-controlled trials. CAF = Coronally Advanced Flap; CTG = CAF + Connective Tissue Graft; BM = CAF + Barrier Membrane; EMD = CAF + Enamel Matrix Derivative; ADM = CAF + Acellular Dermal Matrix; PRP = CAF + Platelet-Rich Plasma.

Two applications of Network Meta-analysis to dental sciences were found searching literature.

Tu et al. (2009) performed a Network Meta-analysis to investigate whether enamel matrix derivatives (EMD) in conjunction with other regenerative materials yield better treatment outcomes than EMD alone in the treatment of infrabony defects ≥ 3 mm. Three (3) outcome variables, probing pocket depth (PPD), clinical attachment level (CAL) and intrabony defect depth were evaluated and 28 trials were included in the study. The diagram as showed in Fig.5 represents a typical example of a network of connected trials.

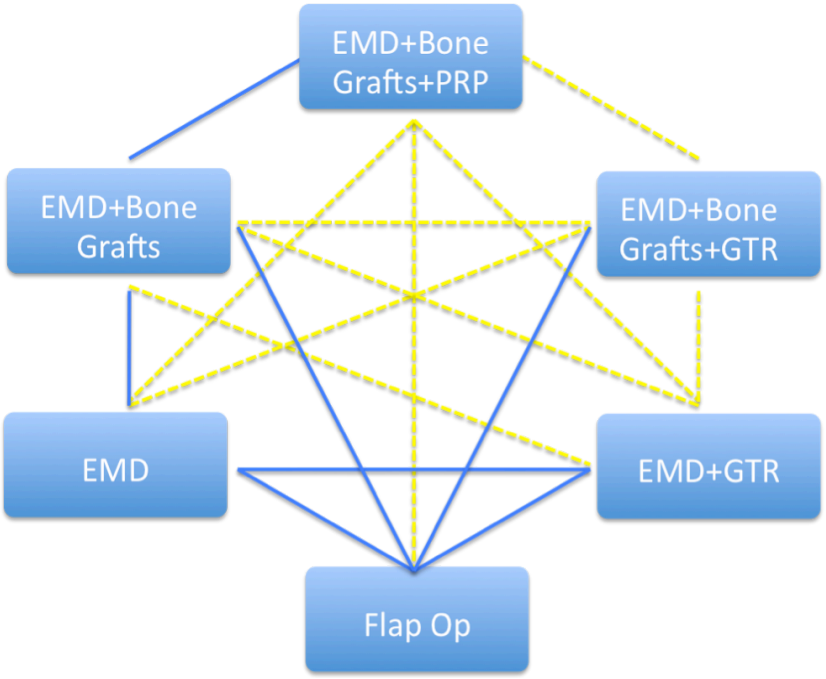


Fig.5 Network for the comparisons among different combination therapies, enamel matrix derivatives (EMD) alone and flap operation. Each node in the network represents a treatment and the solid blue lines between nodes represent pair-wise comparisons for which direct (head-to-head) evidence is available from clinical trials. Dotted yellow lines refer to those comparisons that have not been tested directly in randomized controlled trials.

Network meta-analysis included 27 studies grouped into six nodes. Fiftheen (15) possible pairs of comparisons were present but evidence of direct comparisons was only available in 7 of the 15 pairs. Combining direct and indirect evidence, the authors concluded that there was little evidence to support the additional benefits of EMD in conjunction with other regenerative materials.

A search to determine the relative effectiveness of fluoride toothpastes of different concentrations in preventing dental caries in children and adolescents was undertaken by Walsh et al. (2010). The primary outcome was caries increment in the permanent or deciduous dentition as measured by the change in decayed, (missing), filled tooth surfaces (D(M)FS/d(m)fs) from baseline. Sixty-six (66) studies were included in a Network Meta-analysis where each node of the network diagram represented different concentrations of fluoride toothpastes. The review showed the benefits of using fluoride toothpaste in preventing caries in children and adolescents when compared to placebo, but only significantly for fluoride concentrations of 1000 ppm and above.

Direct and indirect evidence. What do we need to know before combining information?

Combining the data from a set of different studies, indirect comparison and NM approach share all difficulties with standard meta-analysis. Precise definition of treatment procedures, differences in the characteristics of the partecipants, duration of follow-up, outcome measures, quality assessment criteria and others must be accurately considered and subgroup analysis should always be performed. In combining treatments/interventions from different studies, a formal statistical approach should move from a correct specification of assumptions and an appropriate search of trials.

Moreover, researchers should consider some important issues before combining information from direct and indirect evidence (Caldwell 2006):

1. *Are the individual treatments respected or "lumped" together?*

In standard meta-analytic methods, it is often the case that treatments are "lumped" together to form a single comparator allowing for summarizing evidence from different trials. Network Meta-analysis is not constrained like standard meta-analysis and individual treatments are respected and evaluated in the different nodes of the diagram network. For example, the aim of a clinical trial could be to investigate the relative effectiveness of a non-surgical versus a surgical approach to periodontal disease. Surgical treatments of periodontal disease may include Open Flap Debridment, Modified Widman Flap, Regenerative therapy and other techniques. "Lumping" three or more of these surgical treatments to form a single comparator to be compared to non-surgical therapy is questionable and results obtained in this way do not allow for drawing conclusions about which treatment is best on the chosen outcome. Combining direct and indirect evidence, Network Meta-analysis is able to deal with all the possible comparisons and so to avoid problems related to "lumping" treatments.

2. *Do treatments/trials form a connected network?*

Information from direct and indirect evidence can be combined in a Network Meta-analysis only in the case that each treatment/trial is part of a connected network. In other words, in a data-set consisting, for example, of AD, AC, BC, CD, EF, EG, FG pair-wise comparisons, the A, B, C, D group of treatments is not connected to the E, F, G group (Fig.6).

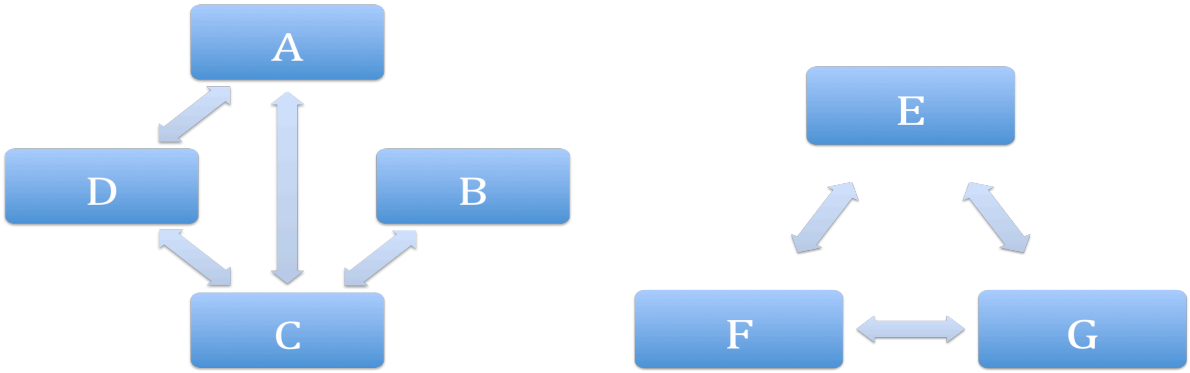


Fig.6 Hypothetical example of a disconnected network of treatment comparisons. Each node in the network represents a treatment and the solid lines between nodes represent pair-wise comparisons for which direct (head-to-head) evidence is available from clinical trials.

3. *Is it appropriate to combine all available evidence in a Network Meta-analysis?*

The key assumption is that the relative effect of one treatment compared with another is the same across the entire set of trials included in the Network meta-analysis. In other words, the relative effect d_{AB} estimated in the A vs B trials should be the same as the $d_{AB} = d_{AC} - d_{BC}$ estimated in the A vs C and B vs C trials (Caldwell 2005). Similarity of trials, inclusion criteria, study partecipants, intervention protocols will support the decision on whether or not to combine direct and indirect evidence. Moreover, clinical experience and informed judgment are required for a careful examination of the trials to be included in the analysis. As with a standard pair-wise meta-analysis, subgroup analysis may be carried out.

Standard meta-analysis - adjusted indirect comparison - Network Meta-analysis. Which approach should be to prefer?

In healthcare decision-making, it is often the case that the clinician has to assess the best treatment strategy according to the relative effectiveness of interventions for a specific condition or patology. Several different treatment options can be available for the same case and it is hard to think that literature could provide direct evidence on all possible treatment comparisons in any of the trials dealing with the same clinical condition.

The relationship between trials can be arbitrarily complex, but conventional meta-analysis can only take into account the results from comparisons of two treatments directly compared and is not able to draw conclusions from the

indirect evidence. This generates a “selection bias”, in terms of automatically excluding evidence with an indirect relationship with the target. Moreover, meta-analysis fails to control the bias which means that it can not produce reliable results if it includes trials of poor methodological quality or badly designed.

Standard meta-analityc techniques, based on pair-wise treatment comparisons, start showing their limitations when dealing with a large network of comparisons. In fact, as the number of treatment choices grows up, the number of possible comparisons of each treatment to another will enhance in a factorial way (Fig.5). In particular, the number of combinations of n treatments taken r at a time is: $C(n,r) = n! / r! (n-r)!$. Thus, considering pair-wise comparisons of treatments, in presence of 4 treatments, researcher will have to deal with 6 comparisons; in presence of 6 treatments, with 15 comparisons; in presence of 10 treatments, with 45 comparisons.

For instance, consider the structures showed in Fig.7, where the letters represent different treatment options, the solid lines represent direct, head-to-head, comparisons between two treatments, and the dotted lines represent indirect comparisons. In this figure, (1) shows a pair-wise treatment comparison, (2) shows an indirect comparison, (3) shows a realistic complex network of comparisons.

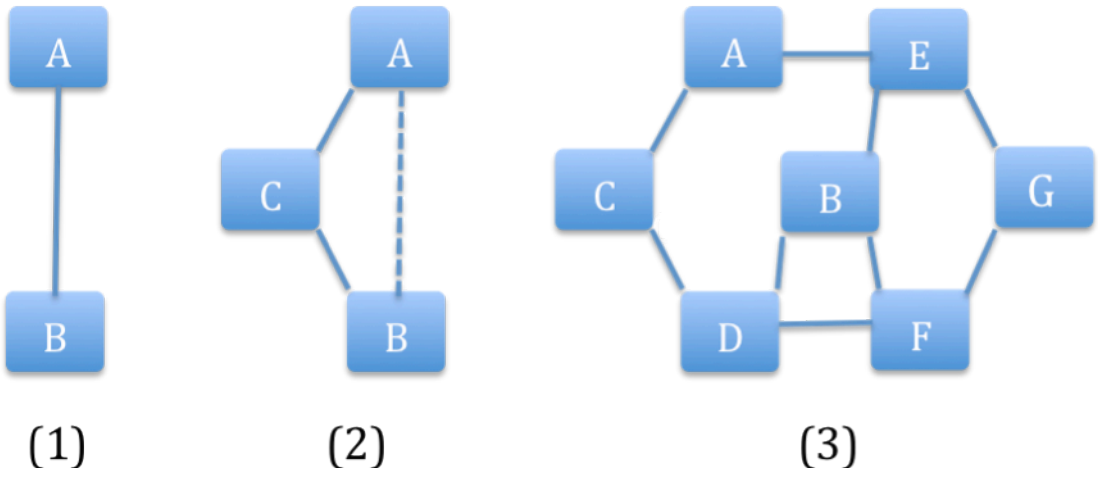


Fig.7 Network evidence structures.

In presence of two treatments to be compared (1), no indirect evidence will be obviously take into account and a single meta-analysis of RCTs providing head-to-head comparison of treatments will represent the most reliable way to summarize evidence of the relative effectiveness of interventions.

In presence of three treatment choices (2) for the same condition, direct evidence will probably be present in literature and strongly raccomended when compared to indirect evidence. However, adjusted indirect comparison methods can be applied in presence of a common comparator and may be particulary useful in demonstrating, for example, that a new active drug is equivalent to (i.e. not very different in efficacy from) an already available active drug, which itself has been shown to be superior to placebo. (Glenny 2005, Hauck 1999, Mainland 1938)

Limitations of adjusted indirect comparison methods can be described as follows:

- adjusted indirect comparison methods do not explicitly generalise to allow inclusion of direct evidence.

- adjusted indirect comparison methods can not take account of trials with three or more arms without either splitting or discarding groups.

In this context, Network Meta-analysis approach provides a method to create a unique diagram of the evidence present in literature, allowing for a large number of treatments to be included and all available evidence from both direct and indirect comparisons to be simultaneously evaluated. Summarizing direct and indirect evidence and avoiding treatments to be grouped in multiple analyses or even excluded, NM approach can represent a really useful way of obtaing an evidence synthesis and a basis in decision making process.

Heterogeneity and Inconsistency

Bringing together material from different studies, one of the most troublesome aspects of many systematic reviews and meta-analysis is quantifying the extent of heterogeneity (Higgins 2002, Thompson 1999, Lu 2004) among a collection of studies.

There are at least three sources of data uncertainty in the Network Meta-analysis data: sampling error, between-trial heterogeneity, Inconsistency (‘Incoherence’) of the network. In particular, heterogeneity and Inconsistency can be described as follows:

- *Between-trial heterogeneity*: between trial heterogeneity within pair-wise comparisons, measured in a similar way as in standard pair-wise meta-analysis. Statistical heterogeneity exists when the true effects being evaluated differ between studies, and may be detectable if the variation between the results of the studies is above that expected by chance (Higgins 2002, Edwards 2009).

- *Inconsistency ('Incoherence') of the network*: between pair-wise comparison heterogeneity. Lumley (2002) suggests comparing the results derived from the different comparators, and suggests a parameter to measure the 'incoherence' of the system, which considers the consistency of a specific estimated contrast between two treatments with the rest of the system.
The authors described an incoherent network of comparisons as showed in Fig.8 (which represents an extreme exemple), where the sign "-" indicates a negative different in treatment effect as two nodes are compared in the direction of the arrow and the sign "+" a positive effect.

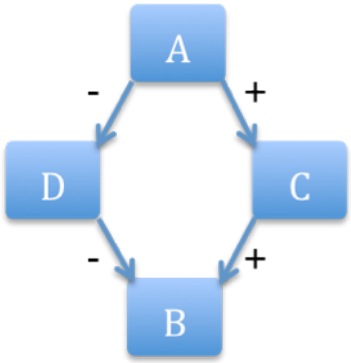


Fig.8 Incoherent network of comparisons.

Combining d_{AC} and d_{CB} suggests that A is better than B, but combining d_{AD} and d_{DB} suggests that A is worse. This evenience may be explained in almost three ways:

1. Estimates are consistent among trials, but these ones are underpowered;
2. Estimates are consistent, but heterogeneity in the effect of one or more treatments is sufficient to account for disagreement in the results.
3. Each individual estimate of d_{AB} is apparently reliable but they disagree. This form of uncertainty can not be handled by standard meta-analytic methods and is called "Inconsistency" (or Incoherence) by Lumley (2002).

The Inconsistency of a network can be due to genuine diversity, bias or combination of both (Salanti 2008). The lack of a demonstrable Inconsistency does not prove that the results are free of bias and diversity. A proper evaluation of Inconsistency may be difficult in epidemiological and clinical studies, due to uncorrect reporting of some important characteristics, presence of few results in many comparisons or results reported in diverse formats.

The Network Meta-analysis approach proposed by some authors such as Lumley (2002) and Lu (2004, 2006) allows for estimating both heterogeneity in the effect of different treatments and Inconsistency in the evidence from different pairs of treatments.

Software for Network Meta-analysis

The most commonly used form of the Network Meta-analysis (or Mixed Treatment Comparison) is based on a Bayesian Markov Chain Monte Carlo method (Spiegelhalter 2002b). Freely available WinBUGS software (Winbugs) offers the greatest flexibility for fitting models. Frequentist approaches have been developed by Lumley (2002) in R using linear mixed models and are also feasible in SAS using PROC NLMIXED (Glenny 2005).

Advantages and limitations of Network Meta-analysis

Meta-analysis of randomized controlled trials with direct treatment comparisons generally provides the most reliable evidence. Network Meta-analysis can be performed in the absence of head-to-head evidence and provides a way to combine direct and indirect comparisons in the same analysis. However, the use of indirect comparison methods and the results of the analysis must be properly interpreted. The unadjusted indirect comparison approach should be avoided (Garthlener 2008) as high frequency of discrepancies from the direct estimate has been showed (Glenny 2005). Adjusted indirect comparison methods and more complex strategies such as Network Meta-analysis can offer a framework for integrating information of arms from multiple trials within a unified analysis of a network of randomized controlled studies. Higgins and Whitehead (1996) showed that the results from direct comparisons could have the precision of their results enhanced by

being analysed with any available indirect comparison of the same treatments (Edwards 2009). However, statistical experience is required to appropriately approach to assumptions underlying the NM model. Sources of evidence have to provide consistent information about the treatments contrasts and further research must be performed to investigate how to interpret results that differ substantially between direct and indirect evidence.

References

About NICE. http://www.nice.org.uk/aboutnice/ (accessed January 2008).

Ades AE. (2003) A chain of evidence with mixed comparisons: models for multiparameter synthesis and consistency of evidence. Stat Med 22: 2995-3016.

Baker SG, Kramer BS. (2002) The transitive fallacy for randomized trials: If A bests B and B bests C in separate trials, is A better than C? BMC Med Res Methodol 2:13.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. (2009) Introduction to meta-analysis. John Wiley & Sons, Ldt.

Bridle C, Bagnall A, Duffy S, et al. (2003) A rapid and systematic review of the clinical and cost-effectiveness of newer drugs for treatment of mania associated with bipolar affective disorder. York, United Kingdom: University of York (http://www.gserve.nice.org.uk/nicemedia/pdf/Bipolar/Assessment//report.pdf).

Bucher HC, Guyatt GH, Griffith LE, et al. (1997) The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 50:683-691.

Cairo F, Pagliaro U, Nieri M. (2008) Treatment of gingival recession with coronally advanced flap procedures: a systematic review. J Clin Periodontol Sep;35(8 Suppl):136-62. Review.

Caldwell DM, Ades AE, Higgins JP. (2005) Simultaneous comparison of multiple treatments: Combining direct and indirect evidence. BMJ 331:897-900.

Caldwell DM. (2006) Introduction to indirect and mixed treatment comparisons. (MTC). Multiple treatment comparisons workshop for NICE. September 1st 2006. http://www.bris.ac.uk/cobm/docs/mtc-indirect.pdf

Eddy DM, Hasselblad V, Shachter R. (1990) A Bayesian method for synthesizing evidence – the Confidence Profile Method. Intern J Technol Assess Health Care 1990; 6: 31–55.Technol Assess Health Care 6: 31-55.

Eddy DM, Hasselblad V, Shachter R. (1992) Meta-analysis by the Confidence Profile Method. London: Academic Press.

Edwards SJ, Clarke MJ, Wordsworth S, Borrill J. (2009) Indirect comparisons of treatments based on systematic reviews of randomised controlled trials. Int J Clin Pract Jun;63(6):841-54.

Egger M, Davey-Smith G, Altman DG. (2001) Systematic Reviews in Health Care: Meta-Analysis in Context. London, United Kingdom: BMJ Publishing Group.

Gartlehner G, Moore CG. (2008) Direct versus indirect comparisons: a summary of the evidence. Int J Technol Assess Health Care Spring;24(2):170-7. Review

Glass GV. (1976) Primary, secondary, and meta-analysis. Edu Res 5:3-8.

Glenny AM, Altman DG, Song F, et al. (2005) Indirect comparisons of competing interventions. Health Technol Assess 9:1-148.14.

Gleser LJ, Olkin I. (1994) Stochastically dependent effect sizes. In: Cooper H, Hedges LV, eds. The handbook of research synthesis. New York: Russell Sage Foundation 339-355.

Hauck WW, Anderson S. (1999) Some issues in the design and analysis of equivalence trials. Drug Information Journal 33:109-18.

Higgins JPT, Whitehead J. (1996) Borrowing strength from external trials in a meta-analysis. Statistics in Medicine 15:2733-2749.

Higgins JPT, Thompson SG. (2002) Quantifying heterogeneity in a meta-analysis. Statist. Med 21:1539-1558.

Higgins JPT, Green S, eds. (2005) Cochrane handbook for systematic reviews of interventions. 4.2.5 [updated May 2005]. In: The Cochrane Library, Issue 3, 2005. Chichester, UK: John Wiley & Sons, Ltd.

Hirotsu C, Yamada Y. (1999) Estimating odds ratios through the connected comparative experiments. Communications in Statistics - Theory and Methods 28:905-29.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. PLoS Med 6(7): e1000100. doi:10.1371/journal.pmed.1000100.

Lu G, Ades AE. (2004) Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 23:3105-3124.

Lu G, Ades AE. (2006) Assessing evidence inconsistency in mixed treatment comparisons. Journal of the American Statistical Association 101: 447-59.

Lumley T. (2002) Network meta-analysis for indirect treatment comparisons. Stat Med 21:2313-2324.

Mainland D. (1938) The treatment of clinical and laboratory data. Edinburgh: Oliver and Boyd. pp. 151-3.

Newhouse PA, Krishnan KR, Doraiswamy PM, et al. (2000) A double-blind comparison of sertraline and fluoxetine in depressed elderly outpatients. J Clin Psychiatry 61:559-568.

Nieri M, Glenny AM, Worthington Helen, Esposito M. (2009) How to interpret meta-analyses of randomized controller trials. Eur J Oral Implantol 2:61-66.

O'Regan C, Ghement I, Eyawo O, Guyatt GH, Mills EJ. (2009) Incorporating multiple interventions in meta-analysis: an evaluation of the mixed treatment comparison with the adjusted indirect comparison. Trials Sep 21;10:86.

Orme M, Collins S, Dakin H, Kelly S, Loftus J. (2010) Mixed treatment comparison and meta-regression of the efficacy and safety of prostaglandin analogues and comparators for primary open-angle glaucoma and ocular hypertension. Curr Med Res Opin Mar;26(3):511-28. Review.

Pearson K. (1904) Report on certain enteric fever inoculation statistics. Br Med J 3:1243-1246.

Phillips A. (2003) Trial and error: cross-trial comparisons of antiretroviral regimens. AIDS 17:619-23.

Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH, et al. (2003) Health outcomes associated with various antihypertensive therapies used as first-line agents: a network meta-analysis. JAMA 289: 2534-44.

Salanti G, Higgins JP, Ades AE, Ioannidis JP (2008) Evaluation of networks of randomized trials. Stat Methods Med Res 17: 279-301.

Song F, Altman DG, Glenny AM, et al. (2003) Validity of indirect comparison for estimating efficacy of competing interventions: Empirical evidence from published meta-analyses. BMJ 326:472.

Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. (2009) Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ Apr 3;338:b1147. doi: 10.1136/bmj.b1147. Review.

Spiegelhalter DJ, Thomas A, Best NG, et al. (2002b) WinBUGS User Manual, Version 1.4. Cambridge, UK: MRC Biostatistics Unit, 2002.

Spiegelhalter DJ, Abrams KR, Myles J. (2004) Bayesian approaches to clinical trials and Health-Care Evaluation. New York: Wiley.

Thompson SG, Sharp SJ. (1999) Explaining heterogeneity in meta-analysis: a comparison of methods. Statist. Med 18:2693-2708.

Tsong Y, Wang SJ, Hung HM, Cui L. (2003) Statistical issues on objective, design, and analysis of noninferiority active-controlled clinical trial. J Biopharm Stat 13:29-41.

Tu YK, Woolston A, Faggion CM Jr. (2009) Do bone grafts or barrier membranes provide additional treatment effects for infrabony lesions treated with enamel matrix derivatives? A network meta-analysis of randomized-controlled trials. J Clin Periodontol. 2010 Jan;37(1):59-79. Epub 2009 Dec 1.

Walsh T, Worthington HV, Glenny AM, Appelbe P, Marinho VCC, Shi X. (2010) Fluoride toothpastes of different concentrations for preventing dental caries in children and adolescents. Cochrane Database of Systematic Reviews Issue 1. Art. No.: CD007868. DOI: 10.1002/14651858.CD007868.pub2.

Wang SJ, Hung HM, Tsong Y. (2002) Utility and pitfalls of some statistical methods in active controlled clinical trials. Control Clin Trials 23:15-28.

Welton NJ, Caldwell DM, Adamopoulos E, Vedhara K. (2009) Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. Am J Epidemiol. 2009 May 1;169(9):1158-65. Epub 2009 Mar 3.

Wilby J, Kainth A, Hawkins N, et al. (2005) A rapid and systematic review of the clinical effectiveness, tolerability and cost effectiveness of newer drugs for epilepsy in adults. Health Technol Assess 9(15):1-172.

WinBUGS with DoodleBUGS. Imperial College and the Medical Research Council, 2007. http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml (accessed January 2008).

Yazdanpanah Y, Sissoko D, Egger M, Mouton Y, Zwahlen M, Chene G. (2004) Clinical efficacy of antiretroviral combination therapy based on protease inhibitors or non-nucleoside analogue reverse transcriptase inhibitors: indirect comparison of controlled trials. BMJ 328:249-53.

2 Network Meta-analysis

2.1 Introduction

Network Meta-analysis (also called Mixed Treatment Comparison - MTC) is an extension of standard meta-analysis. Instead of dealing only with direct comparisons of treatments (as extracted from single trials) and summarizing data in separate meta-analyses, NM approach is able to consider a number of different interventions in a connected network. Our research focused on evaluating several different treatment effects considered in a number of randomized controlled trials in a specific disease. A unique framework including both direct and indirect evidence was developed and all possible pair-wise comparisons of treatments were analyzed at the same time. From a clinical point a view, NM models allowed for obtaining results (and draw conclusions) on treatments never directly compared in single RCTs and ranking the different treatment options.

NM methods were recently proposed by some authors but are still underresearched even if different models were developed in the last years. Throught the search of the literature, we identified two of the most important approches to NM, which were the one proposed by Lu and Ades (2004, 2006), also called Mixed Treatment Comparison - MTC, and the method introduced by Lumley (2002), who described a maximum-likelihood approach using linear mixed models; his method has been applied in relatively few papers (van der Valk et al. 2009, Elliott and Meyer 2007, Psaty et al. 2003).

Following Lu and Ades framework, two models were performed:

- Consistency model (or model under evidence consistency);
- Inconsistency model (or model under evidence inconsistency).

A third model was proposed as a Bayesian extension of the method of Lumley (which was originally based on a frequentist approach).

Finally, in the same spirit as the work of Lumley, we introduced Kullback-Leibler distance as a newly way to evaluate the discrepancy between direct evidence (as analyzed in standard meta-analysis) and information derived by NM models where both direct and indirect evidence are included.

The main issue of these proposals was the ability of assessing different sources of variability and, in particular, to focus on Inconsistency which can

be viewed as a sort of discrepancy that lies between the pairwise comparisons rather than between individual trials (such as heterogeneity).

The objective of our analyses was to explore and compare different approaches to Inconsistency as the literature revealed lack of applications of NM models in which all the sources of variabilty were properly evaluated in combining evidence from different trials.

A Bayesian framework was preferred because of the great flexibilty in dealing with all the sources of uncertainty. The Bayesian analysis produced posterior distributions and measures for individual parameters, such as the posterior mean, median and the relative credible intervals (CrI - e.g., the endpoints of the 90% CrI are the 5 and 95 percentiles of the posterior distribution), as well as posterior distributions for functions of parameters (e.g., estimates of the probability that each treatment is best).

Some NM models can accommodate data from multi-arm trials, but in this work a two-arm trials framework was developed for all the analyses. In presence of trials including more than two arm (so called multi-arm trials) the complexity of the network require different assumptions which are not discussed in this thesis (Higgins 1996, Hasselblad 1998, Domenici 1999, Whitehead 2002).

2.2 Sources of variability

As introduced in chapter 1 of this thesis, at least three sources of data uncertainty can be detected in the Network meta-analysis data: sampling error, heterogeneity and Inconsistency. In particular:

Heterogeneity

Statistical heterogeneity quantify the between-trials variation and exists when the true effects being evaluated differ between studies, and may be detectable if the variation between the results of the studies is above that expected by chance (Higgins 2002).
Ordinary meta-analyses (especially by frequentist methods) customarily evaluate heterogeneity of effects, as a basis for choosing between a fixed-effects or a random-effects procedure. This preliminary step can form part of a Network Meta-analysis, but it should be preceded by examination of potential effect modifiers, because disparities among studies may preclude analysis of the network, even if effects show no heterogeneity within direct comparisons. Deeks et al. (2008) discuss strategies for addressing heterogeneity. The Bayesian approach takes appropriately into account the uncertainty around the heterogeneity variance, while the frequentist approach only uses the point estimate of the heterogeneity variance. A posterior distribution of τ^2 is obtained, making heterogeneity evaluation and investigation more reliable.

Inconsistency

In standard meta-analysis of randomised trials it is assumed that different trials are sufficiently (not necessarily completely) homogeneous and that they estimate the same single treatment effect (fixed effect model) or different treatment effects distributed around a typical value (random effects model). We refer to this assumption for standard meta-analysis as the homogeneity assumption, to distinguish it from other related assumptions.
In Network Meta-analysis, Inconsistency (Incoherence) of the Network can be described as the uncertainty due to discrepancy between direct and indirect inference on pair-wise comparisons. In other words, a departure from consistency arises when the direct and indirect estimates of an effect differ (e.g., the direct estimate d_{BC} does not equal the indirect estimate d_{AC} - d_{AB}). Methods for evaluating Inconsistency have been proposed and represent an important area of research.
In this thesis two of the most important approaches to Inconsistency in NM models was applied and described in the section below.

2.3 Model Specification

We present the sequence of Bayesian hierarchical random-effects models used in this work:

NM Consistency model - Lu and Ades

The framework is based on the statistical model of Higgins and Whitehead (2005).

The formal model is:

$$Y_{jbk} \sim N(d_{bk} + u_j, \sigma_{jbk}{}^2) \qquad j = study$$
$$u_j \sim N(0, \tau^2) \qquad k,b = treatment$$
$$d_{bk} = \mu_k - \mu_b = d_{Ak} - d_{Ab}$$
$$d_{AA} = 0$$

With priors:

$$d_{Ak} \sim normal(0, 10^6)$$
$$\tau \sim uniform[0, 10]$$

where Y_{jbk} = treatment difference estimate from the randomized trial *j* comparing treatments *b* and *k*;

d_{bk} = difference between the true average effects of treatments *k* and *b* so $d_{bk} = \mu_k - \mu_b$ which (expressed in terms of basic parameters) is = $d_{Ak} - d_{Ab}$ (see below) where treatment *k* = *A* is the overall baseline treatment (i.e. A = CAF in the case study presented after in this thesis);

u_j = random effects with variance τ^2 represent the difference between the average effects of treatments *k* and *b* and their effects in this study *j*. They capture heterogeneity of treatment effect. The model assumes the same random-effect variance τ^2 for all treatment comparisons (a fixed-effect model results if $\tau^2 = 0$);

σ_{jbk} = estimated standard error of Y_{jbk} .

When there is no extra information about the parameters besides the available data, the prior densities can be specified by vague prior distributions: *N(0, 10^6)* for the d_{Ak} (independently) and *Uniform(0, 10)* for τ. These priors are common choices in homogeneity models (Lu 2004, 2006, Lumley 2002).

Basic and Functional Parameters

In a Consistency model, the difference (d_{bk}) between the true average effects of treatments *k* and *b* can be expressed in terms of so called 'basic' parameters and 'functional' parameters. The d_{bk} are identified by expressing them in terms of effects relative to treatment A: $d_{bk} = d_{Ak} - d_{Ab}$ with $d_{AA} = 0$ (the order of the subscripts on d_{bk} is conventional, but counterintuitive) where A is assumed as the primary reference treatment.
In other words, any statistical models built on the relations between basic and functional parameters may be called models under evidence consistency.
For instance, in presence of a network of evidence involving A, B, C and D treatments, six treatment effects can be estimated from the data: d_{AB}, d_{AC}, d_{AD}, d_{BC}, d_{BD} and d_{CD}. These values are mathematically related. In fact, if any three are known (and regarded as 'basic') the remaining three can be derived from them (and named 'functional'). The natural choice is, usually, to have treatment A (no treatment) as the reference (baseline) treatment and the three treatment effects relative to baseline as the 'basic' parameters.

Relative to the four treatments A, B, C and D, as represented in Fig.9, the differences in treatment effects d_{BC}, d_{BD} and d_{CD} can be assumed as functional parameters and so expressed in terms of basic parameters d_{AB}, d_{AC} and d_{AD}:

$$\begin{aligned} d_{BC} &= d_{AC} - d_{AB} \\ d_{BD} &= d_{AD} - d_{AB} \\ d_{CD} &= d_{AD} - d_{AC} \end{aligned} \qquad (1)$$

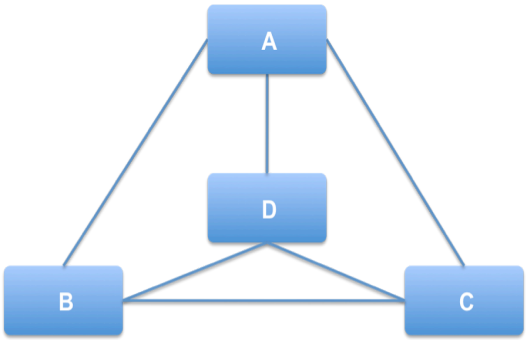


Fig.9 Connected network of comparison of treatment A, B, C, D.

Each relation corresponds to a cycle of edges in the graph.

In a Bayesian framework, it is the basic parameters that must be given prior distributions as follows:

$$d_{AB} \sim normal(0,10^6)$$
$$d_{AC} \sim normal(0,10^6)$$
$$d_{AD} \sim normal(0,10^6)$$

Choices of prior distributions are, to some extent, arbitrary, so they are often subjected to sensitivity analysis. Lambert et al. (2005) discuss sensitivity analysis for exploring the impact of the use of vague priors. On the other hand, some frequentist methods involve approximations and assumptions that are not stated explicitly or verified when the methods are applied. Therefore, both insight into the sensitivity of results from a Bayesian analysis to assumptions on priors and transparent reporting of assumptions underlying a frequentist analysis are highly important.

Any subset of effect parameters can be chosen as basic parameters as long as their corresponding edges can form a spanning tree in the graph *G*, that is, a connected subgraph consisting of all vertices but containing no cycles.

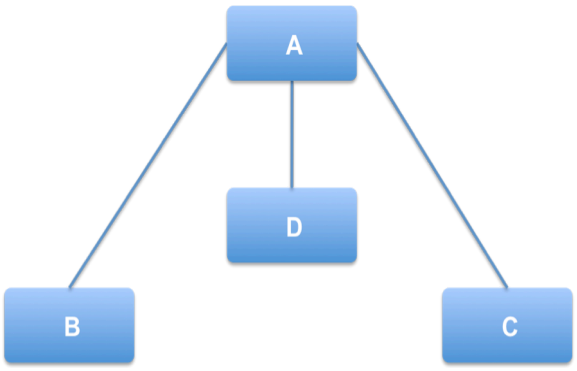


Fig.10 Spanning tree

NM Inconsistency model - Lu and Ades

Following Lumley (2002), Lu and Ades (2006) define evidence Inconsistency as a sort of discrepancy that lies between the pairwise comparisons rather than between individual trials (such as heterogeneity).

At the same time, the way of assessing Inconsistency in network diagrams by these authors is quite different from the model proposed by Lumley (and described after in this chapter). Lu and Ades define Inconsistency as a property of evidence cycles, rather than of individual treatment comparisons.

The authors focused on the structure of networks and expanded the Consistency model by adding one parameter (Inconsistency Factor - ICF or w-factor) for each independent inconsistency.
If the equations in (1) express consistency relations, by definition, inconsistency can be expressed as follows:

$$d_{BC} = d_{AC} - d_{AB} + w_{ABC}$$
$$d_{BD} = d_{AD} - d_{AB} + w_{ABD}$$
$$d_{CD} = d_{AD} - d_{AC} + w_{ACD}$$

where the w-factors, called ICFs, represent the discrepancy between the evidence supporting the functional parameter on the left side and the difference between the basic parameters on the right side. Because each consistency relationship lies on a cycle, the ICFs are, in fact, attached to corresponding cycles rather than to individual edges.
The potential number of inconsistencies in the structure is called by the authors the Inconsistency Degree of Freedom (ICDF). Obviously, the greater the ICDF, the more complex the network structure should be.
For two-arm trials, the ICDF can be calculated as follows:

ICDF = #Functional Parameters = DC – K +1 (for 2-arm trials)

where *DC* (Direct Comparison) is the number treatment comparison informed by data and *K* is the number of treatments included in the network.

Assuming that *ICFs ~ N(0,* τ_w^2*)* then the posterior mean of each ICF is a measure of the extent of inconsistency within the corresponding evidence cycle.

By comparing the models with and without those parameters, one can assess overall Inconsistency, and the posterior distributions of the added parameters

show the extent of inconsistency in the various loops.

Overall Inconsistency (τ_w^2) can be compared with the between-trials heterogeneity τ^2 and expressed in terms of probability as follows:

$Pr\ (\tau_w^2 > \tau^2)$

The posterior probability Pr may give an approximate summary, with a high value signaling potential evidence inconsistency.

A non-informative prior distribution for τ_w^2 is added to the model as follows:

$$\tau_w \sim uniform[0,10]$$

NM Inconsistency model - Bayesian extension of Lumley's method

For networks of two-arm trials that contain loops, Lumley (2002) describe a second random effect w_{kb}, other than heterogeneity, which represents a change in the effect of treatment *k* when it is compared with treatment *b*. In order to combine different treatment comparisons the effect of treatment *k* should be the same no matter what it is compared against, that is, w_{kb} should be close to zero. Thus w_{kb} captures the inconsistency of this pair of treatments with the rest of the evidence. The author uses a frequentist model with one variance parameter to summarize inconsistency (or incoherence) in the network as a whole: τ_w^2 = *var [w]*.

The formal model is thus:

$$Y_{jbk} \sim N(d_{bk} + u_j + w_{bk}, \sigma_{jbk}^2)$$
$$u_j \sim N(0, \tau^2)$$
$$w_{bk} \sim N(0, \tau_w^2)$$

$j = study$

$k, b = treatment$

where Y_{jbk} = treatment difference estimate from the randomized trial *j* comparing treatments *b* and *k*.

u_j = random effects with variance τ^2 represent the difference between the average effects of treatments *k* and *b* and their effects in this study *j*. They capture heterogeneity of treatment effect. The model assumes the same random-effect variance τ^2 for all treatment comparisons. (A fixed-effect model results if $\tau^2 = 0$.)

σ_{jbk} = estimated standard error of Y_{jbk} .

The author proposed a frequentist approach by maximum likelihood or restricted maximum likelihood (REML) estimation in their model (using software for linear mixed models) with one variance parameter to summarize inconsistency (or incoherence) in the network as a whole: τ_w^2 = *var [w]*.

In this thesis a Bayesian approach was used by setting non-informative vague priors as follows:

$$\tau \sim uniform[0,10]$$
$$\tau_w \sim uniform[0,10]$$

Kullback-Leibler method

The Kullback-Leibler distance (Kullback 1959) between two distributions $f_0(t)$ and $f1(t)$, taking $f_0(t)$ as reference, is given by:

$$KL_{f_0}(f_0, f_1) = \int \log \frac{f_1(t)}{f_0(t)} f_0(t) dt$$

To obtain a simple approximation of the Kullback–Leibler distance, we supposed that the posterior distribution of the overall effect was Normal. Let f_0 and f_1 represent the posterior distributions of β obtained, respectively, by the reference and the alternative model. We calculated:

$$KL_{f_0}(f_0, f_1) = 0.5 \times (\frac{(m_o - m_1)^2}{s_1^2} + \frac{s_0^2}{s_1^2} - \log \frac{s_0^2}{s_1^2} - 1)$$

where m_k and s_k^2 are, respectively, the mean and the variance of f_k $(k=0; 1)$.
To better appreciate the amount of KL_{f0} (f_0, f_1) a calibration method was applied (McCulloch 1989). Let $B(p)$ represent the Bernoulli distribution with parameter p. Given a calculated distance d, we calibrated it by the probability q such that:

$$d = KL_{B(0.5)}(B(0.5), B(q))$$
$$0.5 \le q \le 1$$

It can be shown that:

$$q = 0.5 + 0.5(1 - \exp^{-2d})^{0.5}$$

Values of q around 0.5 correspond to low sensitivity of inference, and values close to 1 correspond to substantial changes in results.

In this work, Kullback-Leibler distance (and the correspondent probabilty q) was calculated between the posterior distribution of the mean for the pair-wise comparison under the standard Bayesian meta-analysis model (assumed as the reference) and the posterior distribution of the mean under the NM model (Consistency model without ICFs/w-factors), limited to the subset of treatement comparisons informed by data.

Other approaches to Inconsistency

In a hierarchical Bayesian setting Dias et al. (2010) extended the approach of Bucher et al. (1997) to general networks (but not using evidence from multi-arm trials), by deriving a weighted difference between the estimate from the network and the direct estimate. By plotting the posterior densities of the direct, indirect, and network estimates, they show how the direct evidence and the indirect evidence are combined in the network estimate. For each effect that has direct evidence Dias et al. (2010) also split the information into direct and indirect information and examine the posterior distribution of the difference between the resulting direct and indirect estimates. They discuss how to handle multi-arm trials in this analysis.

Assessment of model fit

In frequentist analyses, assessment of model fit is similar to that for direct evidence and in relation with the particular outcome measure.
For Bayesian model selection and comparison, we used the deviance information criterion *(DIC) DIC* = $D + p_D$ (Spiegelhalter 2002a), where D is the sum of residual deviance and p_D is an estimate of the effective number of parameters, particularly useful in Bayesian hierarchical modeling.

Software

WinBUGS software (Spiegelhalter 2002b) was used for the Bayesian analyses in the case study presented in chapter 3.
The results for SM models are based on a WinBUGS run of 20,000 updates after a 6,000-run burn-in.
The results for NM models are based on a WinBUGS run of 100,000 updates after a 50,000-run burn-in. The method of Gelman and Rubin (1992) was used to check convergence.
Details and programme code are provided in the Appendix A, B and C.

Sensitivity analysis

In hierarchical models it is the choice of prior distribution for the hierarchical variance parameters that has been shown to be most crucial. Thus, in this research, sensitivity analyses focused on evaluating the impact of choosing different prior distributions on τ^2 and τ_w^2 (the areas of greatest uncertainty).
All the alternative priors (other then *Uniform(0, 10)*) for heterogeneity and inconsistency variance which were tested did not produce substantial changes in the values of τ^2 and τ_w^2.

Comparison of the mean residual deviance for individual data points (trials) between models with and without inconsistency factors was also performed. In presence of outliers another Inconsistency model based on the data remaining after outliers trials had been removed was carried out to investigate ICFs and inconsistency variance.

3 Case Study

3.1 Bayesian Network Meta-analysis. Application to surgical treatments of gingival recession

Background

The treatment of buccal gingival recession is a common requirement due to aesthetic concern or root sensitivity in patients with high standards of oral hygiene (American Academy of Periodontology 1996). The ultimate goal of a root coverage procedure is the complete coverage of the recession defect with good appearance related to adjacent soft tissues and minimal probing depth (PD) (Miller 1985, Roccuzzo et al. 2002, Clauser et al. 2003). Previous systematic reviews showed that several surgical procedures such as pedicle flaps, free soft tissue grafts, combinations of pedicle flaps and grafts or barrier membranes (BM) may be indicated to improve the coronal level of the gingival margin on the root surface (Roccuzzo et al. 2002, Clauser et al. 2003, Oates et al. 2003), even if very limited data for epithelialized free gingival graft and laterally positioned flap are available (Roccuzzo et al. 2002). In addition, no difference between resorbable and non-resorbable barriers in terms of mean root coverage was reported (Roccuzzo et al. 2002) and no clinical benefit following root conditioning was detected (Roccuzzo et al. 2002, Cheng et al. 2007). An earlier European Federation of Periodontology Systematic Review on root coverage (Roccuzzo et al. 2002) reported that complete root coverage (CRC) and mean percentage of root coverage varied considerably between studies comparing the same techniques.
The coronally advanced flap (CAF) procedure is a very common approach for root coverage. This procedure is based on the coronal shift of the soft tissues on the exposed root surface (Allen & Miller 1989, Pini Prato et al. 2000). This approach may be used alone or in combination with soft tissue grafts (Wennstrom & Zucchelli 1996), BM (Pini Prato et al. 1992), enamel matrix derivative (EMD) (Rasperini et al. 2000), acellular dermal matrix (ADM) (Harris 1998), platelet-rich plasma (PRP) (Marx et al. 1998) and living tissue-engineered human fibroblast-derived dermal substitute (HF-DDS) (Wilson et al. 2005).

Aim

Two different purposes are identified:

- the *clinical* purpose of this work is to perform a Bayesian Network Meta-analysis to answer the following question: ‘‘Why performing a Network Meta-analysis model instead of standard single meta-analyses in evaluating the CAF (Coronally Advanced Flap) surgical technique alone or in combination with grafts or specific biomaterials (CTG, BM, EMD, ADM, PRP, HF-DDS) in the treatment of Miller Class I and II localized gingival recessions?”

- the *statistical* purpose is to carry out and compare different Bayesian Network Meta-analysis (NM) models according to recent developments of literature. The concept of inconsistency between direct and indirect evidence is investigated.

Material and Methods

This Network Meta-analysis was conducted on the basis of the studies identified and selected in a systematic review by Cairo et al. 2008.

Participants

Patients with a clinical diagnosis of Miller Class I or II localized gingival recession defect.

Studies (N)

Twenty-five (*N* = 25) randomized-controlled clinical trials (RCTs), including a split-mouth model, of at least 6 months’ duration were considered.

Tab.1 Summary of studies and treatment comparisons included in the analysis.

Study	Comparison
da Silva et al. (2004)	CAF *versus* CAF+CTG
Cortellini et al., unpublished data	
Lins et al. (2003)	CAF *versus* CAF+BM
Amarante et al. (2000), Leknes et al. (2005)	
Modica et al. (2000)	CAF *versus* CAF+EMD
Spahr et al. (2005)	
Del Pizzo et al. (2005)	
Castellanos et al. (2006)	
Pilloni et al. (2006)	
Woodyard et al. (2004)	CAF *versus* CAF+ADM
Côrtes et al. (2004, 2006)	
Huang et al. (2005)	CAF *versus* CAF+PRP
Zucchelli et al. (1998)	CAF+CTG *versus* CAF+BM
Jepsen et al. (1998)	
Trombelli et al. (1998)	
Borghetti et al. (1999)	
Tatakis & Trombelli (2000)	
Romagna-Genon (2001)	
Wang et al. (2001)	
McGuire & Nunn (2003)	CAF+ CTG *versus* CAF+EMD
Aichelmann-Reidy et al. (2001)	CAF+CTG *versus* CAF+ADM
Paolantonio et al. (2002)	
Tal et al. (2002)	
Joly et al. (2007)	
Wilson et al. (2005)	CAF+CTG *versus* CAF+HF-DDS

CAF, coronally advanced flap; CTG, connective tissue graft; BM, barrier membrane; EMD, enamel matrix derivative; ADM, acellular dermal matrix graft; PRP, platelet-rich plasma; HF-DDS, human fibroblast-derived dermal substitute.

Outcome measures

The following outcome measures are considered.

- Primary outcome (dycotomous variable)

  Recession defects that obtained Complete Root Coverage (***CRC***). CRC had to be expressed as the number or the percentage of treated teeth of each considered study arm that achieved total root coverage at the follow-up visit.

- Secondary outcomes (continous variables)

  - Change in gingival recession expressed as recession reduction in mm at follow-up visit (***RecRed***),
  - Change in clinical attachment level (CAL) expressed as CAL gain in millimetres at follow-up visit (***CAL gain***),
  - Change in width of keratinized tissue (KT) expressed as KT gain in millimetres at follow-up visit (***KT gain***)

The analyses were conducted on RecRed and CRC outcome measures.

Treatments (K)

The following surgical procedures for the treatment of single recessions are considered:

- ***CAF*** (Coronally Advanced Flap),
- CAF plus Connective Tissue Graft (***CAF+CTG***),
- CAF plus Barrier Membrane (***CAF+BM***),
- CAF plus Enamel Matrix Derivative (***CAF+EMD***),
- CAF plus Acellular Dermal Matrix (***CAF+ADM***),
- CAF plus Platelet Rich Plasma (***CAF+PRP***),
- CAF plus Human Fibroblast-Derived Dermal Substitute (***CAF+HF-DDS***).

Comparisons informed by data (DC-Direct Comparison)

The following comparisons (number of studies) informed by data between the selected techniques were investigated in standard pair-wise meta-analysis (when more than one study per comparison was available) for RecRed, CAL gain and KT gain:

- CAF versus CAF+CTG (2 studies),
- CAF versus CAF+BM (2 studies),
- CAF versus CAF+EMD (5 studies),
- CAF versus CAF+ADM (2 studies),
- CAF versus CAF+PRP (1 studies),
- CAF+CTG versus CAF+BM (6 studies),
- CAF+CTG versus CAF+ADM (4 studies)

The following comparisons (number of studies) informed by data between the selected techniques were investigated in standard pair-wise meta-analysis (when more than one study per comparison was available) for CRC:

- CAF versus CAF+CTG (2 studies),
- CAF vesrsus CAF+BM (1 study),
- CAF versus CAF+EMD (4 studies),
- CAF versus CAF+ADM (2 studies),
- CAF vesrsus CAF+PRP (1 study),
- CAF+CTG versus CAF+BM (6 studies),
- CAF+CTG versus CAF+EMD (1 studies),
- CAF+CTG versus CAF+ADM (4 studies),
- CAF+CTG versus CAF+HF-DDS (1 studies)

Analysis for RecRed

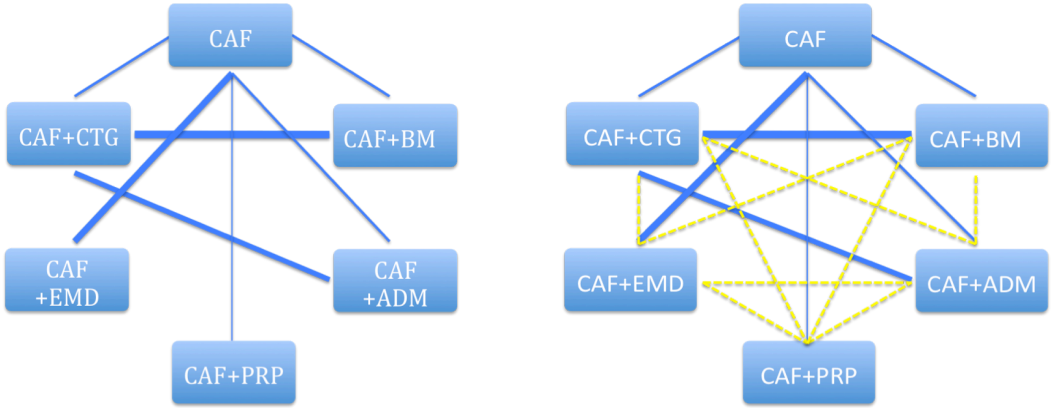
Data

The data (Tab.2) for RecRed outcome consist of:

- *N* = 22 studies
- *K* = 6 treatments (*CAF, CAF+CTG, CAF+BM, CAF+EMD, CAF+ADM, CAF+PRP*)
- *DC* (Direct Comparison) = 7 pair-wise comparisons informed by data (e.g. with direct evidence from single trials)
- *IC* (Indirect Comparison) = 8 pair-wise comparisons based on indirect evidence
- *Comp* = 15 possible pair-wise comparisons (*DC* + *IC*)

Graph Representation

The graphic representation of the network diagram for RecRed data is given in Figure 11.



(a) (b)

Fig.11 Network for the comparisons among 6 different treatments (CAF, CAF+CTG, CAF+BM, CAF+EMD, CAF+ADM, CAF+PRP) without indirect evidence (a), with both direct and indirect evidence. Dotted yellow lines refer to those comparisons that have not been tested directly in randomized-controlled trials. The width of the solid blue lines is in proportion to the amount of evidence (number of RCTs included for each comparison) available in the literature.

Tab.2 NM for RecRed: Summary of studies included in the analysis.

Study	*Treatment Comparison*	*Mean Difference (mm)*	*SD*	*F-up* (months)*	*Study Design*
da Silva et al. (2004)	CAF vs CAF+CTG	0.44	0.27	6	RCT, split mouth design
Cortellini et al. (2008)	CAF vs CAF+CTG	0.52	0.23	6	RCT, parallel study design
Amarante et al. (2000)	CAF vs CAF+BM	-0.20	0.21	6	RCT, split mouth design
Lins et al. (2003)	CAF vs CAF+BM	-0.40	0.29	6	RCT, split mouth design
Modica et al. (2000)	CAF vs CAF+EMD	0.90	0.43	6	RCT, split mouth design
Del Pizzo et al. (2005)	CAF vs CAF+EMD	0.07	0.25	24	RCT, split mouth design
Spahr et al. (2000)	CAF vs CAF+EMD	0.38	0.23	24	RCT, split mouth design
Castellanos et al. (2006)	CAF vs CAF+EMD	0.91	0.44	12	RCT, parallel study design
Pilloni et al. (2006)	CAF vs CAF+EMD	0.93	0.24	18	RCT, parallel study design
Cortes et al. (2006)	CAF vs CAF+ADM	0.08	0.16	24	RCT, split mouth design
Woodyard et al. (2004)	CAF vs CAF+ADM	1.23	0.38	6	RCT, parallel study design
Huang et al. (2005)	CAF vs CAF+PRP	-0.20	0.35	6	RCT, parallel study design
Jepsen et al. (1998)	CAF+CTG vs CAF+BM	-0.01	0.23	12	RCT, split mouth design
Trombelli et al. (1998)	CAF+CTG vs CAF+BM	-0.90	0.26	6	RCT, split mouth design
Zucchelli et al. (1998)	CAF+CTG vs CAF+BM	-0.60	0.21	12	RCT, parallel study design
Borghetti et al. (1999)	CAF+CTG vs CAF+BM	0.00	0.37	6	RCT, split mouth design
Tatakis & Trombelli (2000)	CAF+CTG vs CAF+BM	-0.40	0.24	6	RCT, split mouth design
Wang et al. (2001)	CAF+CTG vs CAF+BM	-0.20	0.27	6	RCT, split mouth design
Aichelmann-Reidy et al. (2001)	CAF+CTG vs CAF+ADM	-0.50	0.29	6	RCT, split mouth design
Paolantonio et al. (2002)	CAF+CTG vs CAF+ADM	-0.20	0.35	12	RCT, parallel study design
Tal et al. (2002)	CAF+CTG vs CAF+ADM	0.29	0.22	12	RCT, split mouth design
Joly et al. (2007)	CAF+CTG vs CAF+ADM	-1.40	0.41	6	RCT, split mouth design

*F-up = Follow-up

Standard pair-wise Bayesian meta-analysis

Standard pair-wise Bayesian meta-analyses for each treatment comparison informed by data were performed when more than one study was present (see also Appendix A) and results are presented in Table 3. Random-effects models were carried out in all cases except when only 2 studies were involved in the analysis and a fixed-effects model was applied to avoid the posterior distribution to include implausibility large values for τ^2. Non informaive prior (the inverse-gamma distribution *(ε, ε)* with *ε = 0.001*) was used in all cases.

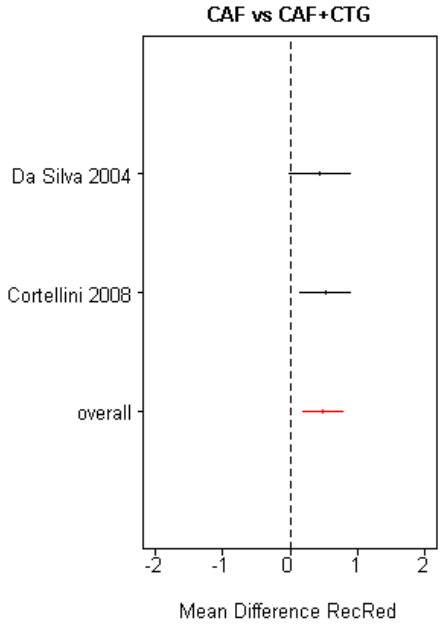
Tab.3 SM for RecRed and relative forest plots.

Standard pair-wise Bayesian meta-analysis – CAF vs CAF+CTG		
Study	***Mean Difference (mm)***	***90% CI****
Da Silva et al. (2000)	0.44	-0.00, 0.88
Cortellini et al. (2005)	0.52	0.14, 0.90
	Point Estimate	***90% CrI****
Overall (FE)	0.49	0.20, 0.78

*90% CI is the 90% Confidence Interval for the Mean Difference.

**Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis.

***90% CrI is the 90% Credible Interval for the Point Estimate.

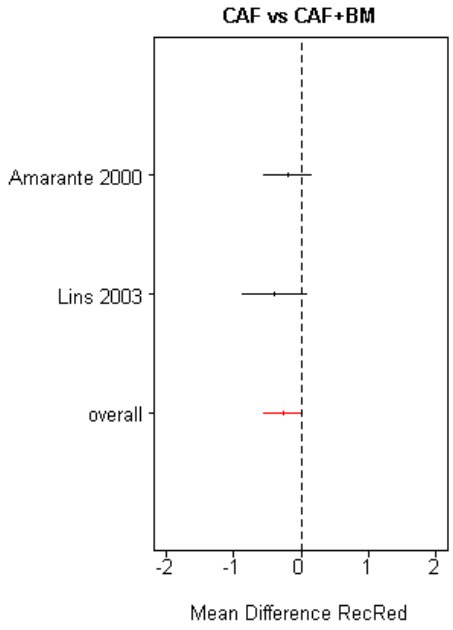


Standard pair-wise Bayesian meta-analysis – CAF vs CAF+BM		
Study	***Mean Difference (mm)***	***90% CI****
Amarante et al. (2000)	-0.20	-0.54, 0.14
Lins et al. (2003)	-0.40	-0.88, 0.08
	Point Estimate*	***90% CrI*****
Overall (FE)	-0.27	-0.55, 0.01

*90% CI is the 90% Confidence Interval for the Mean Difference.

**Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis.

***90% CrI is the 90% Credible Interval for the Point Estimate.

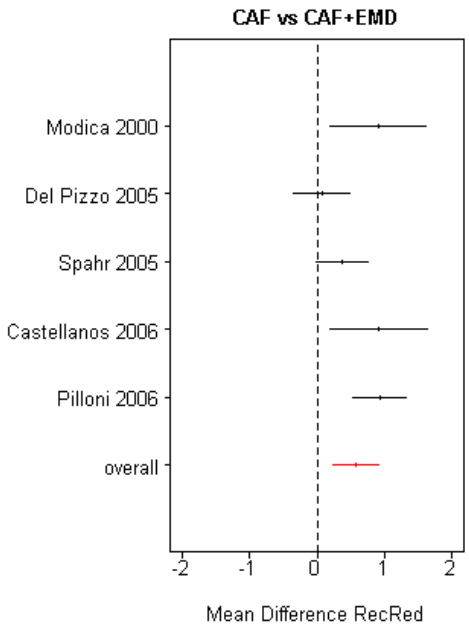


Standard pair-wise Bayesian meta-analysis – CAF vs CAF+EMD		
Study	***Mean Difference (mm)***	***90% CI****
Modica et al. (2004)	0.90	0.20, 1.61
Del Pizzo et al. (2008)	0.07	-0.34, 0.48
Spahr et al. (2000)	0.38	0.00, 0.76
Castellanos et al. (2006)	0.91	0.19, 1.63
Pilloni et al. (2006)	0.93	0.54, 1.32
	Point Estimate*	***90% CrI*****
Overall (RE)	0.57	0.24, 0.91
Tau^2	0.15	0.00, 0.55
I^2	32.26	1.64, 83.24

*90% CI is the 90% Confidence Interval for the Mean Difference.

**Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and I^2 where Point Estimate is the median.

***90% CrI is the 90% Credible Interval for the Point Estimate.

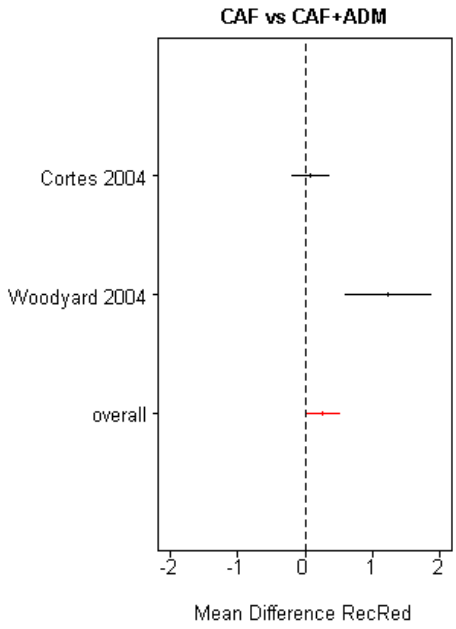


Standard pair-wise Bayesian meta-analysis – CAF vs CAF+ADM		
Study	***Mean Difference (mm)***	***90% CI****
Cortes et al. (2006)	0.08	-0.18, 0.34
Woodyard et al. (2004)	1.23	0.61, 1.85
	Point Estimate*	***90% CrI*****
Overall (FE)	0.25	0.01, 0.50

*90% CI is the 90% Confidence Interval for the Mean Difference.

**Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis.

***90% CrI is the 90% Credible Interval for the Point Estimate.

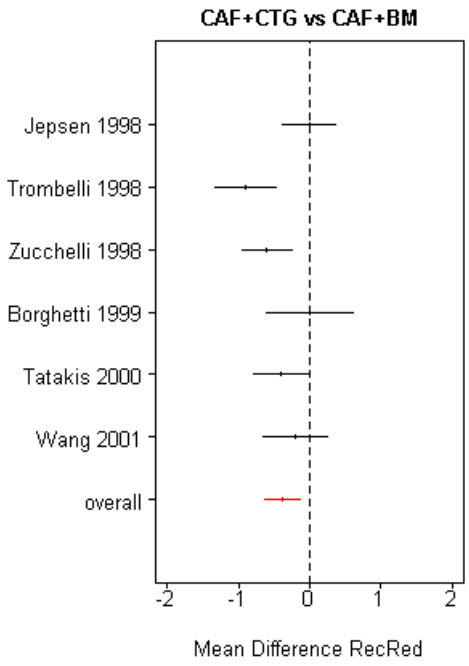


Standard pair-wise Bayesian meta-analysis – CAF+CTG vs CAF+BM		
Study	***Mean Difference (mm)***	***90% CI****
Jepsen et al. (1998)	-0.01	-0.39, 0.37
Trombelli et al. (1998)	-0.90	-1.33, -0.47
Zucchelli et al. (1998)	-0.60	-0.94, -0.26
Borghetti et al. (1999)	0.00	-0.61, 0.61
Tatakis & Trombelli (2000)	-0.40	-0.79, -0.01
Wang et al. (2001)	-0.20	-0.64, 0.24
	Point Estimate*	***90% CrI*****
Overall (RE)	-0.38	-0.62, -0.12
Tau^2	0.08	0.00, 0.30
I^2	32.37	2.17, 80.49

*90% CI is the 90% Confidence Interval for the Mean Difference.

**Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and I^2 where Point Estimate is the median.

***90% CrI is the 90% Credible Interval for the Point Estimate.

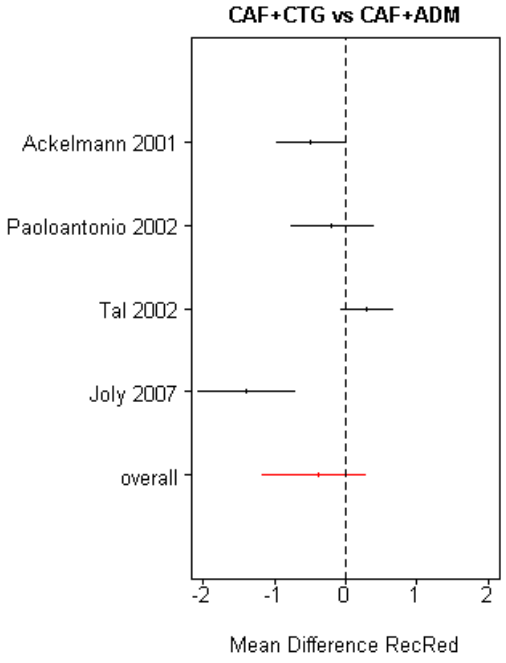


Standard pair-wise Bayesian meta-analysis – CAF+CTG vs CAF+ADM		
Study	***Mean Difference (mm)***	***90% CI****
Aichelmann-Reidy et al. (2001)	-0.50	-0.98, -0.02
Paolantonio et al. (2002)	-0.20	-0.77, 0.37
Tal et al. (2002)	0.29	-0.07, 0.65
Joly et al. (2007)	-1.40	-2.07, -0.73
	Point Estimate	***90% CrI****
Overall (RE)	-0.39	-1.16, 0.28
Tau^2	0.95	0.02, 3.10
I^2	77.47	15.68, 96.70

*90% CI is the 90% Confidence Interval for the Mean Difference.

**Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and I^2 where Point Estimate is the median.

***90% CrI is the 90% Credible Interval for the Point Estimate.



NM Consistency model - Lu and Ades

Among the total $K(K-1)/2 = 15$ potential pairs of comparisons, $DC = 7$ pairwise comparisons are independently supported by direct evidence from the data (Fig. 11). For describing all possible treatment effects in a model, we need to specify $K-1 = 5$ basic parameters that can form a spanning tree. The natural choice si to define the effects of the five combination treatments (i.e., *CAF+CTG, CAF+BM, CAF+EMD, CAF+ADM, CAF+PRP*) relative to the treatment CAF alone.

Therefore, we have the following five basic parameters (represented by the solid blue lines in Fig. 12):

$d_{CAF\ K}$ *where K is CAF+CTG, CAF+BM, CAF+EMD, CAF+ADM, CAF+PRP*

$d_{CAF\ CAF+CTG}$

$d_{CAF\ CAF+BM}$

$d_{CAF\ CAF+EMD}$

$d_{CAF\ CAF+ADM}$

$d_{CAF\ CAF+PRP}$

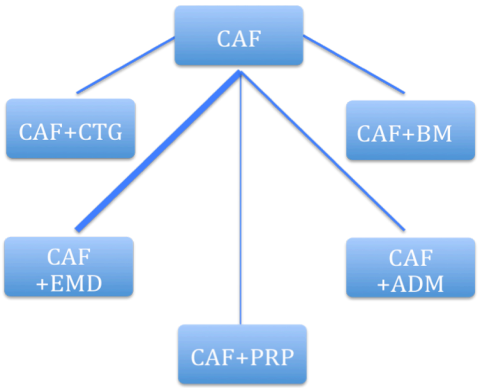


Fig.12 Spanning tree for RecRed

Functional Parameters (on the left of each equation below) can be expressed in terms of difference between basic parameters (on the right of each equation below) (Figure 13) as follows:

$$d_{CAF+CTG\ CAF+BM} = d_{CAF\ CAF+CTG} - d_{CAF\ CAF+BM}$$

$$d_{CAF+CTG\ CAF+ADM} = d_{CAF\ CAF+CTG} - d_{CAF\ CAF+ADM}$$

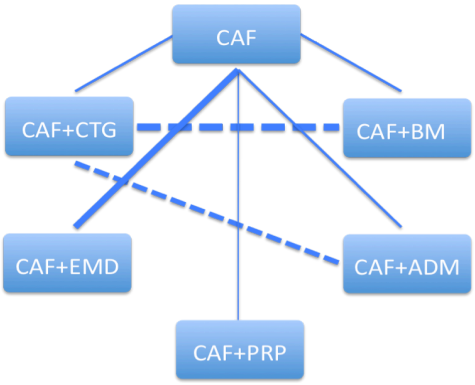


Fig.13 Network of comparisons expressing the relationship between Basic and Functional parameters. The solid lines represent comparisons whose treatment contrasts are specified as basic parameters. All solid lines form a spanning tree. The dotted lines represent comparisons associated with functional parameters.

NM Inconsistency model – Lu and Ades

Inconsistency Degree of Freedom (ICDF)

Potential number of inconsistencies as defined by Lu and Ades (2006) can be calculated as follows in the network diagram for RecRed:

ICDF (for 2-arm trials) = #Functional Parameters = DC – K +1

ICDF for RecRed = #Functional Parameters = 7 – 6 +1 = 2

Two (2) inconsistencies are identified in the NM for RecRed.

Inconsistency Factors (ICFs)/w-factors

Discrepancy between the functional parameters and the difference between the basic parameters can represented as follows:

$$d_{CAF+CTG\ CAF+BM} = d_{CAF\ CAF+CTG} - d_{CAF\ CAF+BM} + W_{CAF\ CAF+CTG\ CAF+BM}$$

$$d_{CAF+CTG\ CAF+ADM} = d_{CAF\ CAF+CTG} - d_{CAF\ CAF+ADM} + W_{CAF\ CAF+CTG\ CAF+ADM}$$

Two (2) ICFs (w-factors) are defined and attached to the corresponding evidence cycles (Figure 14):

$W_{CAF\ CAF+CTG\ CAF+BM}$; $W_{CAF\ CAF+CTG\ CAF+ADM}$

The number of w-factors (2, for RecRed analysis) being estimated is not equal to the number of treatment comparisons (7, for RecRed analysis) (as showed by Lumley model) but is defined according to the presence of functional parameters.

$W_{CAF\ CAF+CTG\ CAF+BM}$ $W_{CAF\ CAF+CTG\ CAF+ADM}$

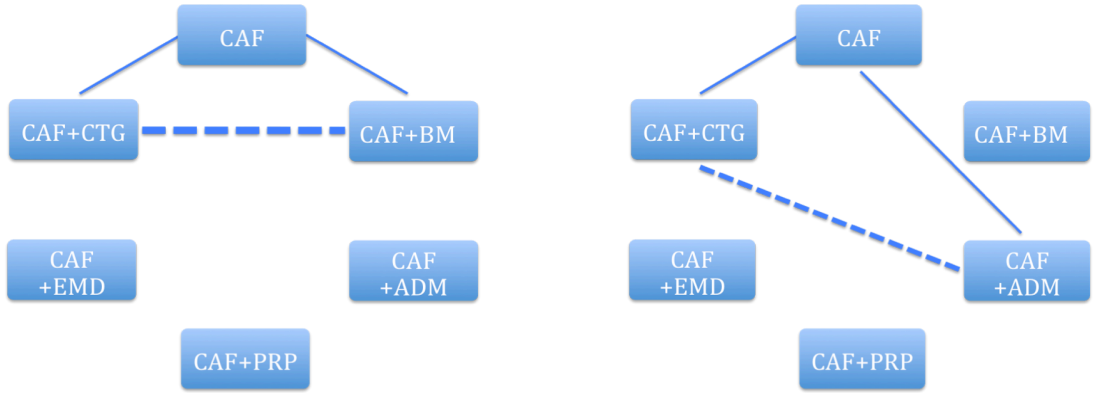


Fig.14 Two (2) evidence cycles are defined by relations between Basic and Functional parameters for the Network Meta-analysis under Consistency model. The solid lines represent comparisons whose treatment contrasts are specified as basic parameters. The dotted lines represent comparisons associated with functional parameters.

Results of NM models with and without Inconsistency Factors

Both in the NM random-effects Consistency and Inconsistency models, the median of the posterior distribution of the between-trials variance (Tau^2) is very small (0.12 and 0.13 respectively), suggesting that similar results would be obtained with a fixed-effects model. Results with and without the inconsistency factors are given in Table 4, 6, 8 (Figure 15, 17).

The goodness-of-fit statistic, *DIC*, for the NM model without ICFs is 25.34 and it is reduced to 24.64 by adding ICFs to the model.

The values of the overall inconsistency, $Tau\text{-}w^2$ = 0.46, and inconsistency probability, $Pr\ (Tau\text{-}w^2 > Tau^2)$ = 0.71, suggest the presence of inconsistency between sources of evidence on posterior treatment effects.

The values of $W_{CAF\ CAF+CTG\ CAF+BM}$ = 0.26, and of $W_{CAF\ CAF+CTG\ CAF+ADM}$ = - 0.22 suggest the presence of inconsistency in the two evidence cycles.

Combinations of CAF+EMD and CAF+CTG showed the best results in terms of RecRed under both Consistency and Inconsistency models and occuped the first and the second position respectively in the ranking of treatments and in terms of probability of beeing the best (Table 5, 7; Figure 16, 18)

Mean residual deviance for individual data points in models with and without inconsistency factors has been compared and Figure 19 shows the presence of two outliers, corresponding to trials 11 and 22, in the model without w-factors. Point 11 correspond to the comparison CAF vs CAF+PRP of the trial of Woodyard et al. 2004 and point 22 to the comparison CAF+CTG vs CAF+ADM of the trial of Joly et al. 2007. The last one confirm suspicions about the (CAF CAF+CTG CAF+ADM) cycle.

Another inconsistency model has been carried out after removing points 11 and 22 from the analysis. The results (Table 9) show both ICFs ($W_{CAF\ CAF+CTG\ CAF+BM}$ = 0.20, $W_{CAF\ CAF+CTG\ CAF+ADM}$ = 0.14) and overall inconsistency ($Tau\text{-}w^2$ = 0.25) are substantially reduced, not the inconsistency probability due to the simoutaneous strong reduction of between-trials variance (Tau^2 = 0.05). Results of Consistency and Inconsistency (full data and 11, 22 deleted) models are showed in Table 10.

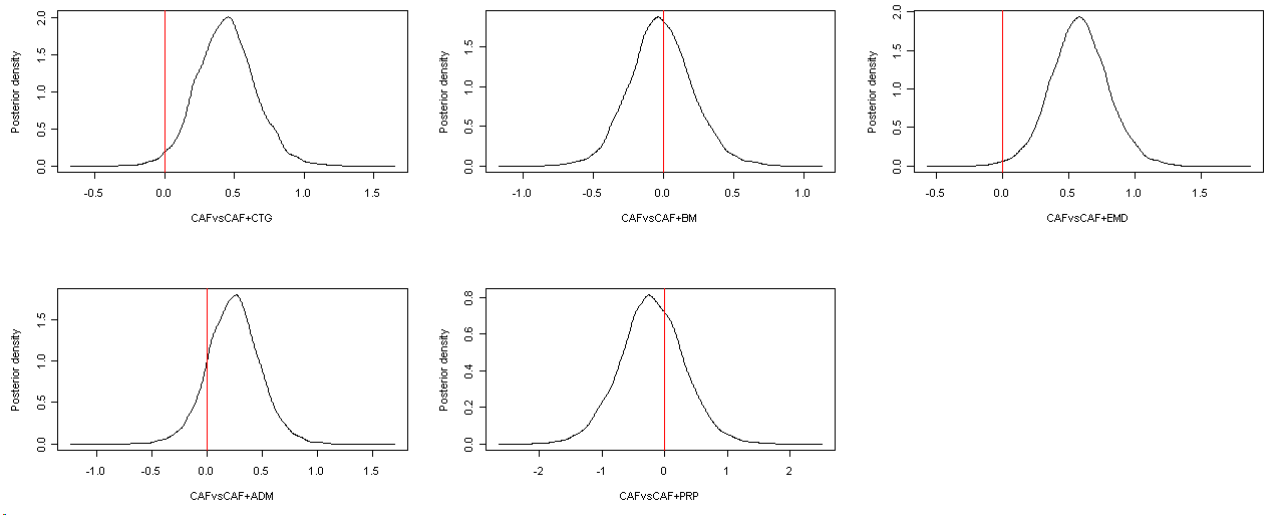
Tab.4 NM for RecRed: Consistency model (without ICFs/w-factors). Results of all possible pair-wise treatment comparisons.

Network Meta-analysis – Consistency model		
Treatment Comparison	***Point Estimate****	***90% CrI****
CAF vs CAF+CTG	0.44	0.11, 0.79
CAF vs CAF+BM	-0.02	-0.38, 0.36
CAF vs CAF+EMD	0.58	0.23, 0.94
CAF vs CAF+ADM	0.24	-0.14, 0.63
CAF vs CAF+PRP	-0.20	-1.06, 0.64
CAF+CTG vs CAF+BM	-0.46	-0.73, -0.18
CAF+CTG vs CAF+EMD	0.14	-0.36, 0.62
CAF+CTG vs CAF+ADM	-0.20	-0.55, 0.14
CAF+CTG vs CAF+PRP	-0.65	-1.56, 0.26
CAF+BM vs CAF+EMD	0.60	0.10, 1.12
CAF+BM vs CAF+ADM	0.26	-0.16, 0.67
CAF+BM vs CAF+PRP	-0.18	-1.11, 0.73
CAF+EMD vs CAF+ADM	-0.34	-0.87, 0.18
CAF+EMD vs CAF+PRP	-0.78	-1.71, 0.13
CAF+ADM vs CAF+PRP	-0.45	-1.37, 0.49
Tau^2	0.12	0.03, 0.33
DIC	25.34	

*Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 where Point Estimate is the median.

**90% CrI is the 90% Credible Interval for the Point Estimate.

Fig.15. NM for RecRed: Plots of the posterior densities of each treatment compared to CAF under Consistency model (without ICFs/w-factors).



Tab.5 NM for RecRed: Ranking and Best for the six treatments included in the analysis under Consistency model (without ICFs/w-factors).

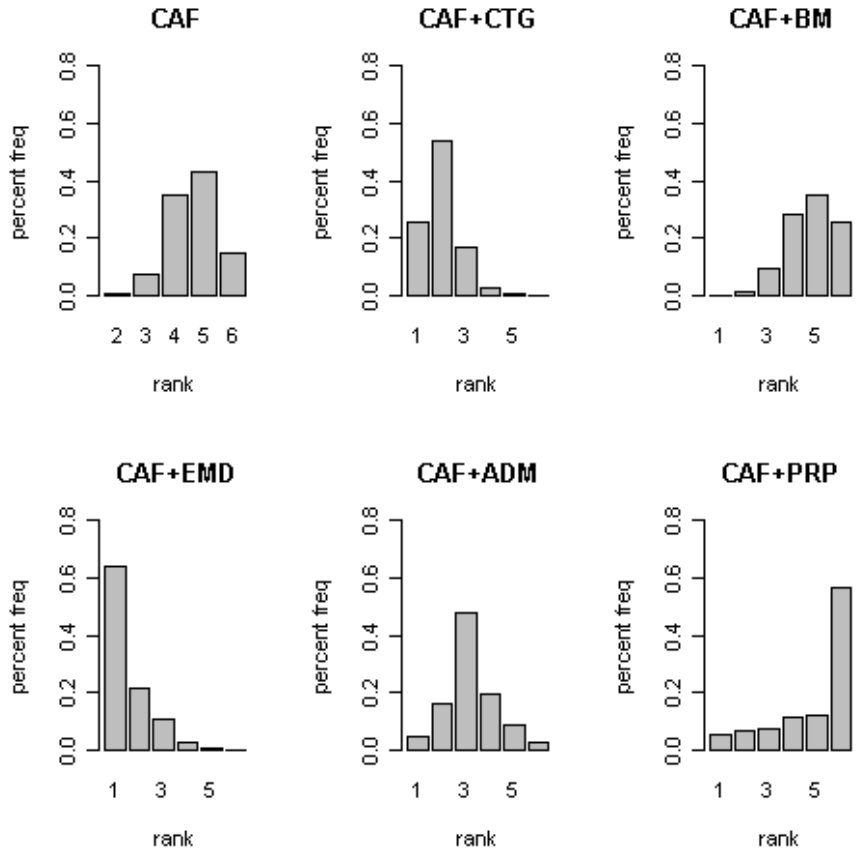
	Ranking		***Best***
Treatment	***Point Estimate****	***90% CrI****	***Pr*****
CAF	4.63	3.00, 6.00	0.00
CAF+CTG	1.99	1.00, 3.00	0.26
CAF+BM	4.75	3.00, 6.00	<0.00
CAF+EMD	1.55	1.00, 3.00	0.64
CAF+ADM	3.19	2.00, 5.00	0.05
CAF+PRP	4.89	1.00, 6.00	0.05

*Point Estimate is the median of the posterior distribution of the Bayesian meta-analysis.

**90% CrI is the 90% Credible Interval for the Point Estimate.

***Pr is the probability that each treatment is the best.

Fig.16 NM for RecRed: Ranking for the six treatments under Consistency model (without ICFs/w-factors).



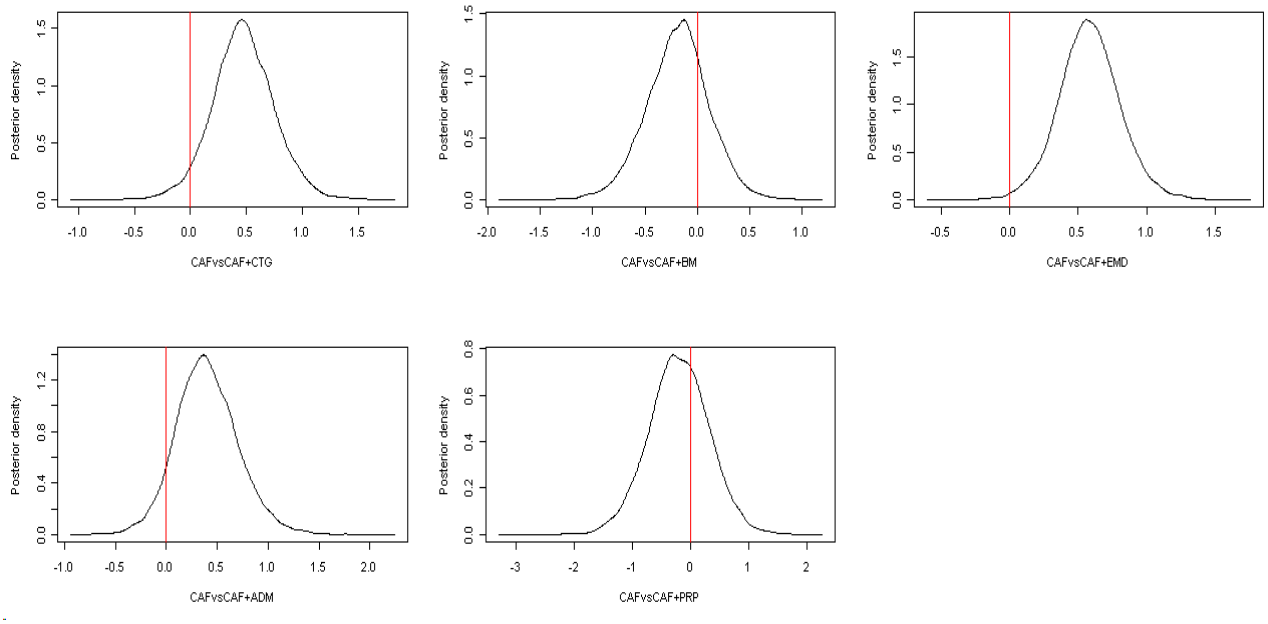
Tab.6 NM for RecRed: Inconsistency model (with ICFs/w-factors). Results of all possible pair-wise treatment comparisons.

Network Meta-analysis – Inconsistency model (with ICFs/w-factors)		
Treatment Comparison	***Point Estimate****	***90% CrI****
CAF vs CAF+CTG	0.47	0.03, 0.94
CAF vs CAF+BM	-0.19	-0.69, 0.29
CAF vs CAF+EMD	0.58	0.22, 0.95
CAF vs CAF+ADM	0.40	-0.08, 0.92
CAF vs CAF+PRP	-0.20	-1.06, 0.65
CAF+CTG vs CAF+BM	-0.66	-1.32, -0.09
CAF+CTG vs CAF+EMD	0.11	-0.47, 0.69
CAF+CTG vs CAF+ADM	-0.07	-0.66, 0.57
CAF+CTG vs CAF+PRP	-0.67	-1.63, 0.28
CAF+BM vs CAF+EMD	0.77	0.18, 1.40
CAF+BM vs CAF+ADM	0.59	-0.07, 1.34
CAF+BM vs CAF+PRP	-0.01	-0.10, 0.98
CAF+EMD vs CAF+ADM	-0.18	-0.79, 0.45
CAF+EMD vs CAF+PRP	-0.78	-1.73, 0.13
CAF+ADM vs CAF+PRP	-0.60	-1.59, 0.37
w-CAF.CTG.BM	0.26	-0.31, 0.98
w-CAF.CTG.ADM	-0.22	-0.96, 0.36
Tau^2	0.13	0.03, 0.34
Tau-w^2	0.46	0.00, 24.38
Pr (Tau-w^2 > Tau^2)	0.71	
DIC	24.64	

**Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and Tau-w^2 where Point Estimate is the median (90% CrI is the 90% Credible Interval).

**90% CrI is the 90% Credible Interval for the Point Estimate.

Fig.17 NM for RecRed: Plots of posterior densities of each treatment compared to CAF under the Inconsistency model (with ICFs/w-factors).



Tab.7 NM for RecRed: Ranking and Best for the six treatments included in the analysis under Inconsistency model (with ICFs/w-factors).

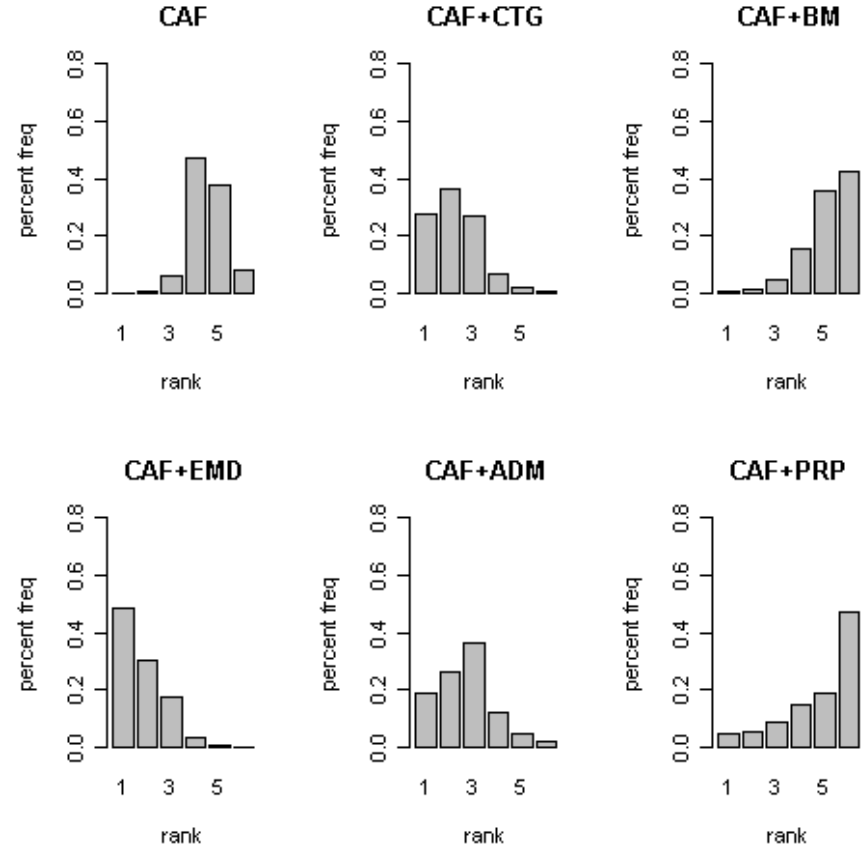
	Ranking		***Best***
Treatment	***Point Estimate****	***90% CrI****	***Pr*****
CAF	4.47	3.00, 6.00	<0.00
CAF+CTG	2.21	1.00, 4.00	0.28
CAF+BM	5.12	3.00, 6.00	0.00
CAF+EMD	1.78	1.00, 3.00	0.48
CAF+ADM	2.62	1.00, 5.00	0.19
CAF+PRP	4.80	2.00, 6.00	0.05

*Point Estimate is the median of the posterior distribution of the Bayesian meta-analysis.

**90% CrI is the 90% Credible Interval for the Point Estimate.

***Pr is the probability that each treatment is the best.

Fig. 18 NM for RecRed: Ranking for the six treatments under Inconsistency model (with ICFs/w-factors).



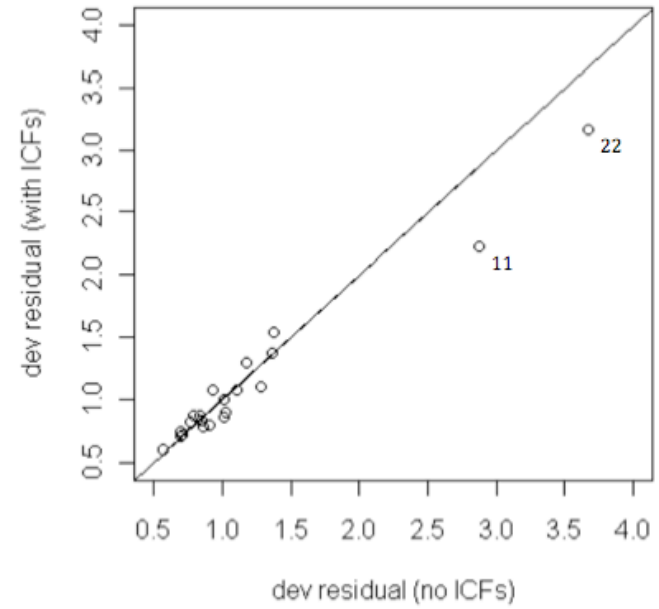
Tab.8 NM for RecRed: Standard meta-analysis, Consistency and Inconsistency models (limited to the comparisons informed by data and analyzed in SM).

	SM	***NM (Consistency model)***	***NM (Inconsistency model)***
Treatment Comparison (N*)	***Point Estimate (90% CrI)*****	***Point Estimate (90% CrI)***	***Point Estimate (90% CrI)***
CAF vs CAF+CTG (2)	0.49 (0.20, 0.78)	0.44 (0.11, 0.79)	0.47 (0.03, 0.94)
CAF vs CAF+BM (2)	-0.27 (-0.55, 0.01)	-0.02 (-0.38, 0.36)	-0.19 (-0.69, 0.29)
CAF vs CAF+EMD (5)	0.57 (0.24, 0.91)	0.58 (0.23, 0.94)	0.58 (0.22, 0.95)
CAF vs CAF+ADM (2)	0.25 (0.01, 0.50)	0.24 (-0.14, 0.63)	0.40 (-0.08, 0.92)
CAF+CTG vs CAF+BM (6)	-0.38 (0.62, 0.12)	-0.46 (-0.73, -0.18)	-0.66 (-1.32, -0.09)
CAF+CTG vs CAF+ADM (4)	-0.39 (-1.16, 0.28)	-0.20 -0.55, 0.14	-0.07 (-0.66, 0.57)
Tau2	-	0.12 (0.03, 0.33)	0.13 (0.03, 0.34)
Tau-w2	-	-	0.46 (0.00, 24.38)
Pr (Tau-w2 > Tau2)	-	-	0.71
DIC	-	25.34	24.64

**Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and $Tau\text{-}w^2$ where Point Estimate is the median (90% CrI is the 90% Credible Interval).

**90% CrI is the 90% Credible Interval for the Point Estimate.

Fig.19 NM for RecRed: residuals in fitting models with and without Inconsistency Factors. Points 11 and 22 correspond to the comparison CAF vs CAF+PRP of the trial of Woodyard et al. 2004 and CAF+CTG vs CAF+ADM of the trial of Joly et al. 2007.



Tab.9 NM for RecRed: Inconsistency model (11 and 22 deleted). Results of all possible pair-wise treatment comparisons.

Network Meta-analysis – Inconsistency model (11 and 22 deleted)		
Treatment Comparison	***Point Estimate****	***90% CrI****
CAF vs CAF+CTG	0.38	0.01, 0.77
CAF vs CAF+BM	-0.21	-0.60, 0.15
CAF vs CAF+EMD	0.57	0.28, 0.87
CAF vs CAF+ADM	0.13	-0.32, 0.57
CAF vs CAF+PRP	-0.19	-0.91, 0.52
CAF+CTG vs CAF+BM	-0.59	-1.14, -0.14
CAF+CTG vs CAF+EMD	0.19	-0.29, 0.67
CAF+CTG vs CAF+ADM	-0.25	-0.81, 0.25
CAF+CTG vs CAF+PRP	-0.57	-1.38, 0.24
CAF+BM vs CAF+EMD	0.78	0.32, 1.28
CAF+BM vs CAF+ADM	0.34	-0.21, 0.91
CAF+BM vs CAF+PRP	0.02	-0.79, 0.83
CAF+EMD vs CAF+ADM	-0.44	-0.99, 0.08
CAF+EMD vs CAF+PRP	-0.76	-1.54, 0.01
CAF+ADM vs CAF+PRP	-0.32	-1.16, 0.53
w-CAF.CTG.BM	0.20	-0.24, 0.76
w-CAF.CTG.ADM	0.14	-0.37, 0.76
Tau^2	0.05	0.00, 0.20
$Tau\text{-}w^2$	0.25	0.00, 16.12
Pr ($Tau\text{-}w^2 > Tau^2$)	0.74	
DIC	17.77	

**Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and $Tau\text{-}w^2$ where Point Estimate is the median (90% CrI is the 90% Credible Interval).

**90% CrI is the 90% Credible Interval for the Point Estimate.

Tab.10 NM for RecRed: Consistency, Inconsistency (full data) and Inconsistency (11 and 22 deleted) models (limited to the comparisons informed by data and analyzed in SM).

		NM (Inconsistency model)	
	NM (Consistency model)	***Full data***	***11 and 22 deleted***
Treatment Comparison (N*)	***Point Estimate (90% CrI)****	***Point Estimate (90% CrI)***	***Point Estimate (90% CrI)***
CAF vs CAF+CTG (2)	0.44 (0.11, 0.79)	0.47 (0.03, 0.94)	0.38 (0.01, 0.77)
CAF vs CAF+BM (2)	-0.02 (-0.38, 0.36)	-0.19 (-0.69, 0.29)	-0.21 (-0.60, 0.15)
CAF vs CAF+EMD (5)	0.58 (0.23, 0.94)	0.58 (0.22, 0.95)	0.57 (0.28, 0.87)
CAF vs CAF+ADM (2)***	0.24 (-0.14, 0.63)	0.40 (-0.08, 0.92)	0.13 (-0.32, 0.57)
CAF+CTG vs CAF+BM (6)	-0.46 (-0.73, -0.18)	-0.66 (-1.32, -0.09)	-0.59 (-1.14, -0.14)
CAF+CTG vs CAF+ADM (4)****	-0.20 -0.55, 0.14	-0.07 (-0.66, 0.57)	-0.25 (-0.81, 0.25)
w-CAF.CTG.BM	-	0.26 (-0.31, 0.98)	0.20 (-0.24, 0.76)
w-CAF.CTG.ADM	-	-0.22 (-0.96, 0.36)	0.14 (-0.37, 0.76)
Tau^2	0.12 (0.03, 0.33)	0.13 (0.03, 0.34)	0.05 (0.00, 0.20)
Tau-w^2	-	0.46 (0.00, 24.38)	0.25 (0.00, 16.12)
Pr (Tau-w^2 > Tau^2)	-	0.71	0.74
DIC	25.34	24.64	17.77

*N is the number of trials with direct evidence for each treatment comparison.

**Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and Tau-w^2 where Point Estimate is the median (90% CrI is the 90% Credible Interval).

***In the Inconsistency model without 11 and 22, N = 1.

****In the Inconsistency model without 11 and 22, N = 3.

NM Consistency model (Heterogenous variance) - Lu and Ades

In Consistency and Inconsistency models proposed, all random effects are assumed having the same variance, which is just the case of homogeneity of between-trial variation (Higgins 1996, Lu 2004). Heterogeneity of between-trial variation has been investigated and an Heterogeneous model has been performed but discarded due to a higher value of DIC with respect to Homogeneous models previously described. Results are presented in Table 11.

Tab.11 NM for RecRed: Heterogeneous model. Results of all possible pair-wise treatment comparisons.

Network Meta-analysis – Heterogeneous model				
Treatment Comparison	***Point Estimate****	***90% CrI****	σ^2	***90% CrI***
CAF vs CAF+CTG	0.40	0.08, 0.72	0.16	0.00, 0.60
CAF vs CAF+BM	-0.06	-0.38, 0.30	0.21	0.00, 0.78
CAF vs CAF+EMD	0.57	0.24, 0.93	0.14	0.00, 0.50
CAF vs CAF+ADM	0.20	-0.28, 0.66	0.55	0.01, 2.06
CAF vs CAF+PRP	-0.19	-1.39, 1.04	1.08	0.00, 3.98
CAF+CTG vs CAF+BM	-0.46	-0.70, -0.22	0.10	0.00, 0.33
CAF+CTG vs CAF+EMD	0.17	-0.30, 0.66	-	-
CAF+CTG vs CAF+ADM	-0.19	-0.62, 0.25	0.40	0.01, 1.38
CAF+CTG vs CAF+PRP	-0.59	-1.82, 0.66	-	-
CAF+BM vs CAF+EMD	0.63	0.13, 1.11	-	-
CAF+BM vs CAF+ADM	0.27	-0.22, 0.74	-	-
CAF+BM vs CAF+PRP	-0.13	-1.39, 1.13	-	-
CAF+EMD vs CAF+ADM	-0.37	-0.97, 0.20	-	-
CAF+EMD vs CAF+PRP	-0.76	-2.02, 0.53	-	-
CAF+ADM vs CAF+PRP	-0.39	-1.69, 0.94	-	-
Tau^2	0.10	0.01, 0.68		
DIC	26.23			

*Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 where Point Estimate is the median.

**90% CrI is the 90% Credible Interval for the Point Estimate.

NM Inconsistency model - Bayesian extension of Lumley's method

Among the total $K(K-1)/2 = 15$ potential pairs of comparisons, $DC = 7$ pairwise comparisons are independently supported by direct evidence from the data (Fig. 1).

The number of terms w_{kb} being estimated is equal to the number of treatment comparisons. Thus, seven (7) w-factors can identified for RecRed analysis:

$w_{CAF\ CAF+CTG}$, $w_{CAF\ CAF+BM}$, $w_{CAF\ CAF+EMD}$, $w_{CAF\ CAF+ADM}$

$w_{CAF\ CAF+PRP}$, $w_{CAF+CTG\ CAF+BM}$, $w_{CAF+CTG\ CAF+ADM}$

Results for the Bayesian extension of Lumley's method

The median of the posterior distribution of the between-trials variance (Tau^2) is very small (0.13) and similar to models previously described. Results for all possible pair-wise comparisons are given in Table 12. Posterior densities of each treatment compared to reference treatment (CAF) are represented in Figure 20.

The goodness-of-fit statistic, *DIC*, for this NM model (24.76) is close o the value of Inconsistency model by Lu and Ades (24.64).

The lower values (with respect to Lu and Ades model) of the overall Inconsistency, $Tau\text{-}w^2 = 0.18$, and Inconsistency probability, $Pr\ (Tau\text{-}w^2 > Tau^2) = 0.57$, may be explained according to the fact that in the present model the w-factors are attached to single comparisons informed by data and not to evidence cycles.

The values of w-factors relative to CAFvsCAF+BM (- 0.12), CAFvsCAF+ADM (- 0.12), CAFvsCAF+PRP (0.17), CAF+CTGvsCAF+BM (0.09), CAF+CTGvsCAF+ADM (- 0.38) and asimmetries in the relative posterior densities (Figure 22) suggest the presence of Inconsistency in correspondence of these comparisons (Table 12).

Combinations of CAF+EMD and CAF+CTG showed the best results in terms of RecRed also under the Bayesian extension of Lumley's method (Table 13, Figure 21).

Mean residual deviance (Figure 23) for individual data points in the Bayesian extension of Lumley's method show similar results to the Inconsistency model by Lu and Ades.

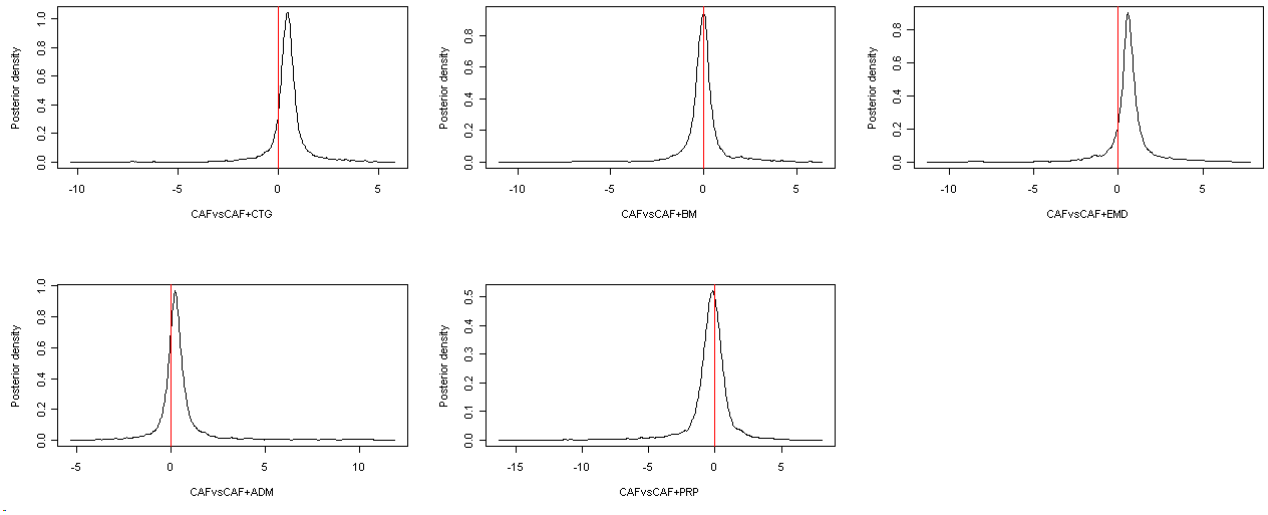
Tab.12 NM for RecRed: Bayesian extension of Lumley's method. Results of all possible pair-wise treatment comparisons.

Network Meta-analysis – Bayesian extension of Lumley's method				
Treatment Comparison	***Point Estimate****	***90% CrI****	***w***	***90% CrI***
CAF vs CAF+CTG	0.43	-0.94, 1.85	0.04	-1.36, 1.41
CAF vs CAF+BM	-0.07	-1.55, 1.74	-0.12	-1.93, 1.22
CAF vs CAF+EMD	0.60	-1.29, 2.65	-0.01	-2.06, 1.84
CAF vs CAF+ADM	0.52	-0.89, 2.60	-0.12	-2.09, 1.35
CAF vs CAF+PRP	-0.37	-2.88, 1.57	0.17	-1.56, 2.57
CAF+CTG vs CAF+BM	-0.50	-1.82, 0.90	0.09	-1.26, 1.44
CAF+CTG vs CAF+EMD	0.17	-2.11, 2.88	-	-
CAF+CTG vs CAF+ADM	0.09	-1.19, 2.28	-0.38	-2.60, 0.83
CAF+CTG vs CAF+PRP	-0.80	-3.42, 1.50	-	-
CAF+BM vs CAF+EMD	0.67	-1.87, 3.41	-	-
CAF+BM vs CAF+ADM	0.59	-0.98, 3.17	-	-
CAF+BM vs CAF+PRP	-0.31	-3.14, 2.08	-	-
CAF+EMD vs CAF+ADM	-0.09	-2.78, 3.45	-	-
CAF+EMD vs CAF+PRP	-0.98	-4.19, 1.64	-	-
CAF+ADM vs CAF+PRP	-0.89	-4.47, 1.39	-	-
Tau^2	0.13	0.04, 0.36	-	-
Tau-w^2	0.18	0.00, 15.15		
Pr (Tau-w^2 > Tau^2)	0.57			
DIC	24.76			

*Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and Tau-w^2 where Point Estimate is the median.

**90% CrI is the 90% Credible Interval for the Point Estimate.

Fig.20 NM for RecRed: Plots of posterior densities of each treatment compared to CAF under the Bayesian extension of Lumley's method.



Tab.13 NM for RecRed: Ranking and Best for the six treatments included in the analysis under the Bayesian extension of Lumley's method.

	Ranking		***Best***
Treatment	***Point Estimate****	***90% CrI****	***Pr*****
CAF	4.27	2.00, 6.00	0.01
CAF+CTG	2.47	1.00, 5.00	0.23
CAF+BM	4.42	2.00, 6.00	0.04
CAF+EMD	2.25	1.00, 6.00	0.48
CAF+ADM	3.08	1.00, 6.00	0.16
CAF+PRP	4.51	1.00, 6.00	0.09

*Point Estimate is the median of the posterior distribution of the Bayesian meta-analysis.

**90% CrI is the 90% Credible Interval for the Point Estimate.

***Pr is the probability that each treatment is the best.

Fig. 21 NM for RecRed: Ranking for the seven treatments under the Bayesian extension of Lumley's method.

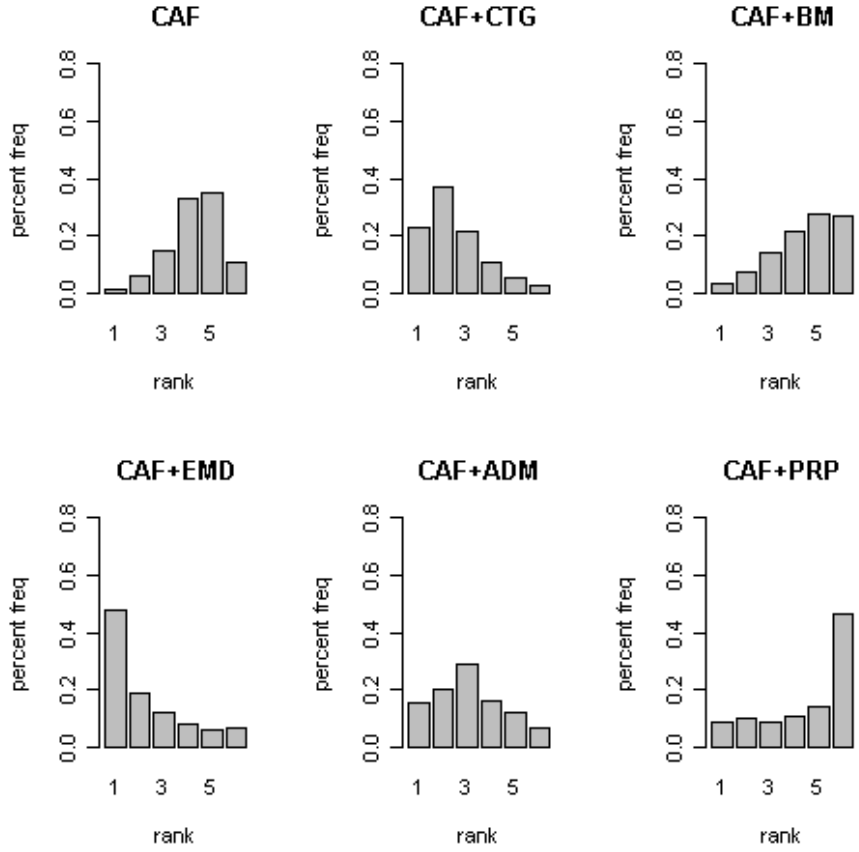


Fig.22 NM for RecRed: Plots of posterior densities of w-factors under the Bayesian extension of Lumley's method.

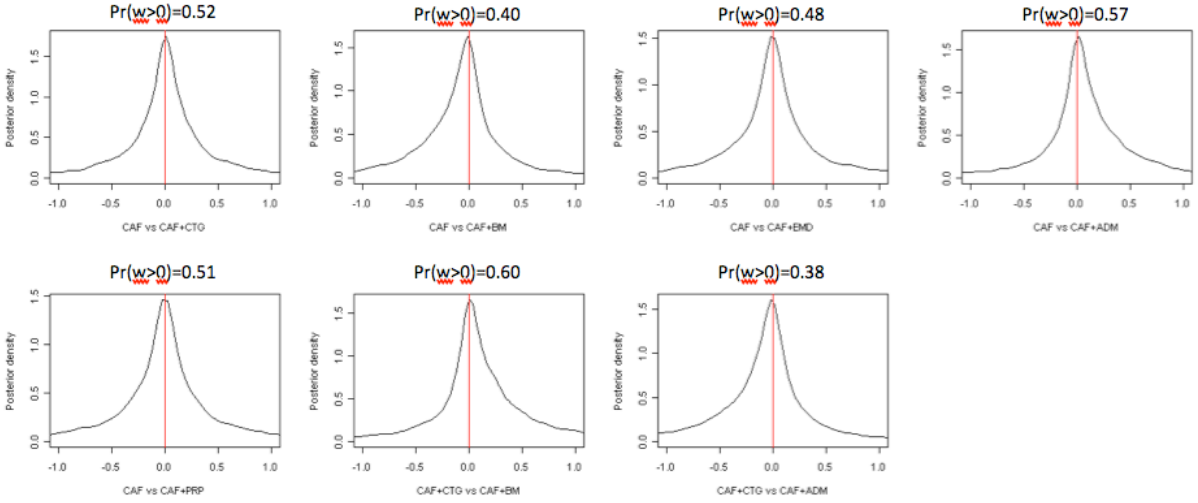
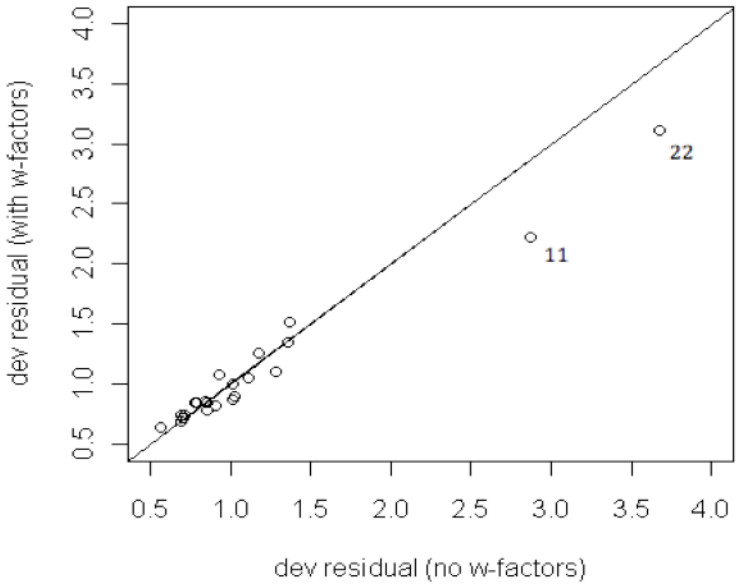


Fig.23 NM for RecRed: residuals in fitting Consistency models without w-factors and the Bayesian extension of Lumley's method. Points 11 and 22 correspond to the comparison CAF vs CAF+PRP of the trial of Woodyard et al. 2004 and CAF+CTG vs CAF+ADM of the trial of Joly et al. 2007.



Kullback Leibler distance

Kullback-Leibler distance between the posterior distribution of treatment effect obtained by SM model (assumed as reference distribution) and by NM Consistency model was calculated for each treatment comparison informed by data (Table 14). KL distance (and relative q) for CAFvsCAF+BM (0.67;0.93), CAFvsCAF+ADM (0.17;0.77), CAF+CTGvsCAF+BM (0.13;0.74), CAF+CTGvsCAF+ADM(1.97;1.00) and corresponding plots (Figure 24) suggest the presence of discrepancy between direct and indirect evidence for these comparisons according to the results of the Bayesian extension of Lumley's method. KL distances (*q*) and w-factors (as derived by Bayesian extension of Lumley's method) are compared in Table 15.

Tab.14 NM for RecRed: Consistency model (without ICFs/w-factors) and standard pair-wise Bayesian meta-analysis (limited to the comparisons informed by data and analyzed in SM).

	Network Meta-analysis (Consistency model)		***Standard pair-wise meta-analysis (Ref.)***		
Treatment Comparison (N*)	***Point Estimate*****	***90% CrI******	***Point Estimate***	***90% CrI***	***KL(q)*******
CAF vs CAF+CTG (2)	0.44	0.11, 0.79	0.49	0.20, 0.78	0.05 (0.66)
CAF vs CAF+BM (2)	-0.02	-0.38, 0.36	-0.27	-0.55, 0.01	0.67 (0.93)
CAF vs CAF+EMD (5)	0.58	0.23, 0.94	0.57	0.24, 0.91	0.00 (0.54)
CAF vs CAF+ADM (2)	0.24	-0.14, 0.63	0.25	0.01, 0.50	0.17 (0.77)
CAF+CTG vs CAF+BM (6)	-0.46	-0.73, -0.18	-0.38	-0.62, -0.12	0.13 (0.74)
CAF+CTG vs CAF+ADM (4)	-0.20	-0.55, 0.14	-0.39	-1.16, 0.28	1.97 (1.00)

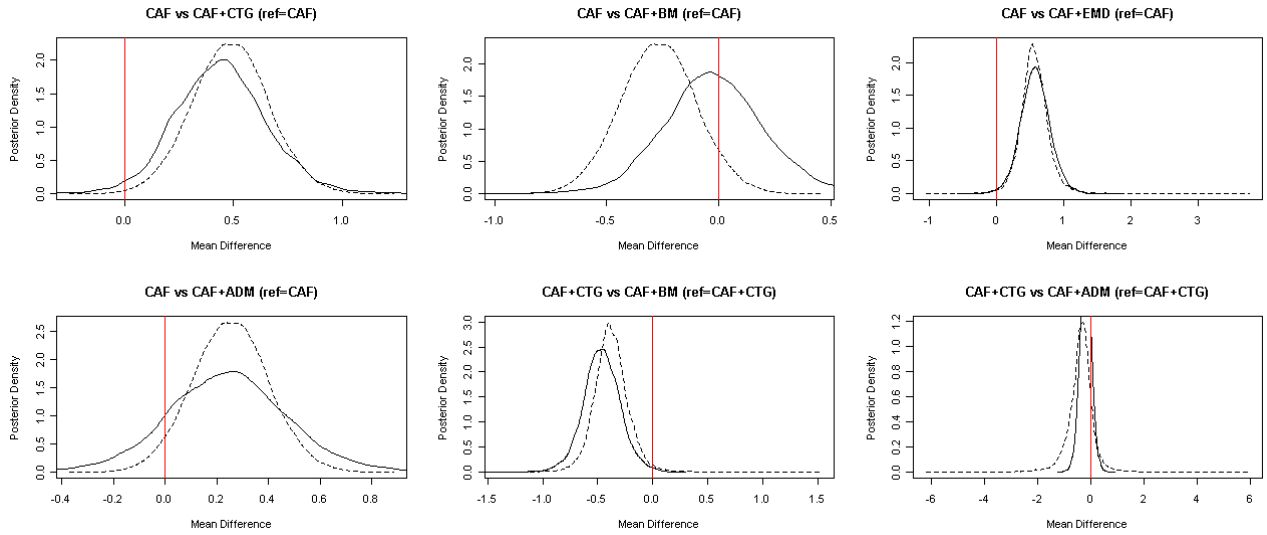
*N is the number of trials with direct evidence for the treatment comparison.

**Point Estimate is the mean of the posterior distribution of the Bayesian meta-analysis.

***90% CrI is the 90% Credible Interval for the Point Estimate.

****KI is the Kullback–Leibler distance between the posterior distribution of the mean for the pair-wise comparison under the NM model and the correspondent pair-wise comparison under the standard meta-analysis model; q is the calibrated value of the Kullback–Leibler distances espressed as a probability.

Fig.24 NM for RecRec: Plots of posterior distributions of Consistency model (without ICFs/w-factors) and standard pair-wise Bayesian meta-analysis (limited to the comparisons informed by data and analyzed in SM).



Dotted lines = Network Meta-analysis; solid lines = Standard pair-wise Bayesian Meta-analysis.

Tab.15 NM for RecRed: SM, NM Consistency (Lu and Ades) and NM (Bayesian extension of Lumley's method) models (limited to the comparisons informed by data and analyzed in SM).

	SM (Ref.)	***NM (Consistency model)***		***NM (Lumley)***	
Treatment Comparison (N*)	***Point Estimate (90% CrI)*****	***Point Estimate (90% CrI)***	***KL(q)******	***Point Estimate (90% CrI)***	***w***
CAF vs CAF+CTG (2)	0.49 (0.20, 0.78)	0.44 (0.11, 0.79)	0.05 (0.66)	0.43 (-0.94, 1.85)	0.04 (-1.36, 1.41)
CAF vs CAF+BM (2)	-0.27 (-0.55, 0.01)	-0.02 (-0.38, 0.36)	0.67 (0.93)	-0.07 (-1.55, 1.74)	-0.12 (-1.93, 1.22)
CAF vs CAF+EMD (5)	0.57 (0.24, 0.91)	0.58 (0.23, 0.94)	0.00 (0.54)	0.60 (-1.29, 2.65)	-0.01 (-2.06, 1.84)
CAF vs CAF+ADM (2)	0.25 (0.01, 0.50)	0.24 (-0.14, 0.63)	0.17 (0.77)	0.52 (-0.89, 2.60)	-0.12 (-2.09, 1.35)
CAF+CTG vs CAF+BM (6)	-0.38 (0.62, 0.12)	-0.46 (-0.73, -0.18)	0.13 (0.74)	-0.50 (-1.82, 0.90)	0.09 (-1.26, 1.44)
CAF+CTG vs CAF+ADM (4)	-0.39 (-1.16, 0.28)	0.20 -0.55, 0.14	1.97 (1.00)	0.09 (-1.19, 2.28)	-0.38 (-2.60, 0.83)
Tau2	-	0.12 (0.03, 0.33)	-	0.13 (0.04, 0.36)	-
Tau-w2	-	-	-	0.18 (0.00, 15.15)	-
Pr (Tau-w2 > Tau2)	-	-	-	0.57	-
DIC	-	25.34	-	24.56	-

*N is the number of trials with direct evidence for each treatment comparison.

**Point Estimate is the mean of the posterior distribution of the bayesian meta-analysis (90% CrI is the 90% Credible Interval).

***KI is the Kullback– Leibler distance between the posterior distribution of the Mean for the pair-wise comparison under the NM model and the correspondent pair-wise comparison under the standard meta-analysis model; q is the calibrated value of the Kullback–Leibler distances espressed as a probability.

Analysis for CRC

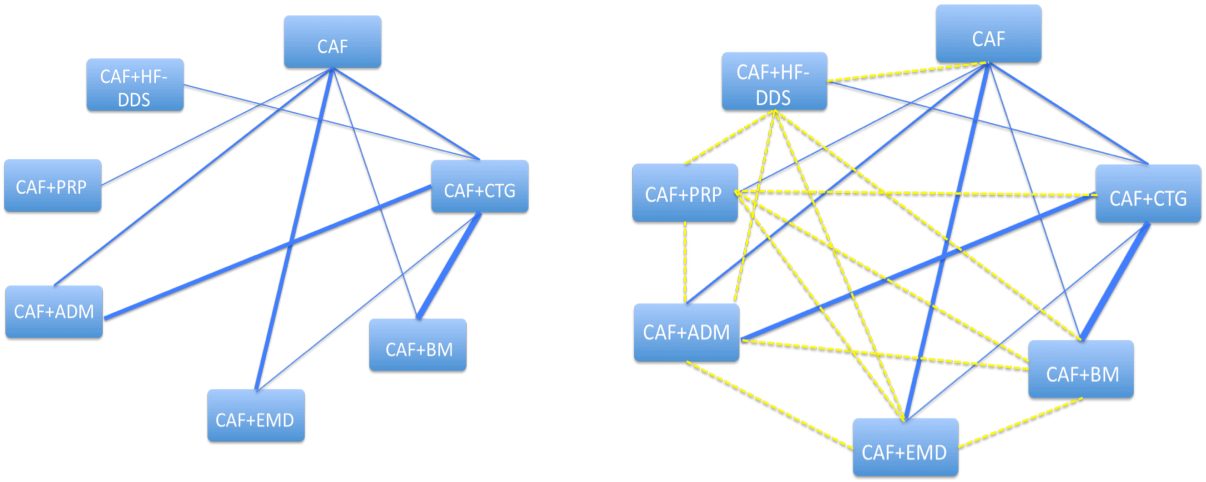
Data

The data (Tab.16) for CRC outcome consist of:

- *N* = 22 studies
- *K* = 7 treatments (*CAF, CAF+CTG, CAF+BM, CAF+EMD, CAF+ADM, CAF+PRP, CAF+HF-DDS*)
- *DC* (Direct Comparison) = 9 pair-wise comparisons informed by data (e.g. with direct evidence from single trials)
- *IC* (Indirect Comparison) = 12 pair-wise comparisons based on indirect evidence
- *Comp* = 21 possible pair-wise comparisons (*DC* + *IC*)

Graph Representation

The graphic representation of the network diagram for CRC data is given in Figure 25.



(a) (b)

Fig.25 Network for the comparisons among 7 different treatments (CAF, CAF+CTG, CAF+BM, CAF+EMD, CAF+ADM, CAF+PRP, CAF+HF-DDS) without indirect evidence (a), with both direct and indirect evidence. Dotted yellow lines refer to those comparisons that have not been tested directly in randomized-controlled trials. The width of the solid blue lines is in proportion to the amount of evidence (number of RCTs included for each comparison) available in the literature.

Tab.16 NM for CRC Summary of studies included in the analysis.

Study	*Treatment Comparison*	*LogOR*	*SD*	*Follow-up (months)*	*Study Design*
da Silva et al. (2004)	CAF vs CAF+CTG	0.80	1.15	6	RCT, split mouth design
Cortellini et al. (2008)	CAF vs CAF+CTG	0.93	0.45	6	RCT, parallel study design
Leknes et al. (2005)	CAF vs CAF+BM	-0.54	0.65	12	RCT, split mouth design
Modica et al. (2000)	CAF vs CAF+EMD	0.00	1.00	6	RCT, split mouth design
Del Pizzo et al. (2005)	CAF vs CAF+EMD	1.61	1.55	24	RCT, split mouth design
Spahr et al. (2000)	CAF vs CAF+EMD	1.32	0.50	24	RCT, split mouth design
Pilloni et al. (2006)	CAF vs CAF+EMD	2.57	0.94	18	RCT, parallel study design
Cortes et al. (2006)	CAF vs CAF+ADM	0.00	1.29	24	RCT, split mouth design
Woodyard et al. (2004)	CAF vs CAF+ADM	3.09	1.21	6	RCT, parallel study design
Huang et al. (2005)	CAF vs CAF+PRP	0.22	0.86	6	RCT, parallel study design
Jepsen et al. (1998)	CAF+CTG vs CAF+BM	0.00	0.82	12	RCT, split mouth design
Zucchelli et al. (1998)	CAF+CTG vs CAF+BM	-1.39	0.61	6	RCT, split mouth design
Borghetti et al. (1999)	CAF+CTG vs CAF+BM	0.00	0.73	12	RCT, parallel study design
Trombelli et al. (1998)	CAF+CTG vs CAF+BM	-2.58	1.08	6	RCT, split mouth design
Tatakis & Trombelli (2000)	CAF+CTG vs CAF+BM	-1.95	1.51	6	RCT, split mouth design
Wang et al. (2001)	CAF+CTG vs CAF+BM	0.00	0.82	6	RCT, split mouth design
McGuire & Nunn (2003)	CAF+CTG vs CAF+EMD	0.84	0.83	12	RCT, split mouth design
Aichelmann-Reidy et al. (2001)	CAF+CTG vs CAF+ADM	-0.76	0.55	6	RCT, split mouth design
Paolantonio et al. (2002)	CAF+CTG vs CAF+ADM	-0.88	0.78	12	RCT, parallel study design
Tal et al. (2002)	CAF+CTG vs CAF+ADM	0.00	0.82	12	RCT, split mouth design
Joly et al. (2007)	CAF+CTG vs CAF+ADM	-2.20	1.49	6	RCT, split mouth design
Wilson Jr (2005)	CAF+CTG vs CAF+HF-DDS	0.00	2.00	6	RCT, split mouth design

Standard pair-wise Bayesian meta-analysis

Standard pair-wise Bayesian meta-analyses for each treatment comparison informed by data were performed when more than one study was present (see also Appendix A) and results are presented in Table 17. Random-effects models were carried out in all cases except when only 2 studies were involved in the analysis and a fixed-effects model was applied to avoid the posterior distribution to include implausibility large values for τ^2. Non informaive prior (the inverse-gamma distribution *(ε, ε)* with *ε = 0.001*) were used in all cases.

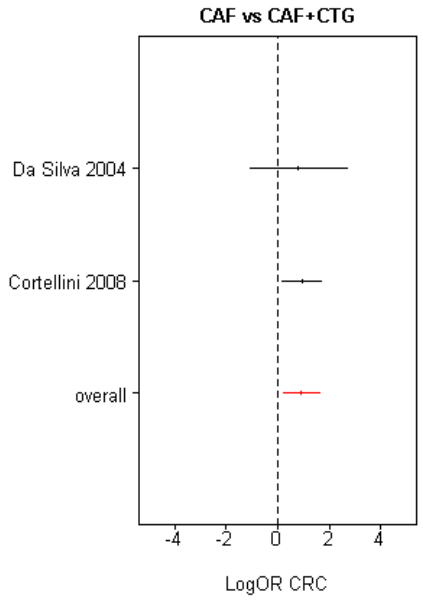
Tab.17 SM for CRC and relative forest plots.

Standard pair-wise Bayesian meta-analysis – CAF vs CAF+CTG		
Study	***LogOR***	***90% CI****
Da Silva et al. (2000)	0.80	-1.09, 2.69
Cortellini et al. (2005)	0.93	0.19, 1.67
	LogOR*	***90% CrI*****
Overall (FE)	0.91	0.23, 1.61

*90% CI is the 90% Confidence Interval for the LogOR.

**LogOR= Log Odds Ratio of the posterior distribution of the Bayesian meta-analysis.

***90% CrI is the 90% Credible Interval for LogOR.

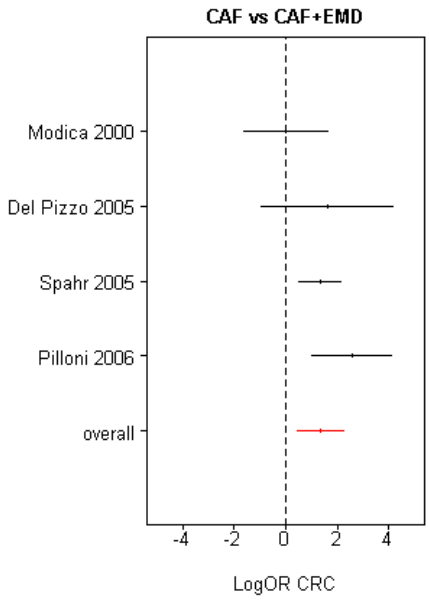


Standard pair-wise Bayesian meta-analysis – CAF vs CAF+EMD		
Study	***LogOR***	***90% CI****
Modica et al. (2004)	0.00	-1.64, 1.64
Del Pizzo et al. (2008)	1.61	-0.93, 4.15
Spahr et al. (2000)	1.32	0.50, 2.14
Pilloni et al. (2006)	2.57	1.03, 4.11
	LogOR*	***90% CrI*****
Overall (RE)	1.36	0.43, 2.25
Tau^2	0.07	0.00, 3.37
I^2	5.59	0.12, 74.84

*90% CI is the 90% Confidence Interval for the LogOR.

**LogOR is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and I^2 where LogOR is the median.

***90% CrI is the 90% Credible Interval for the LogOR.

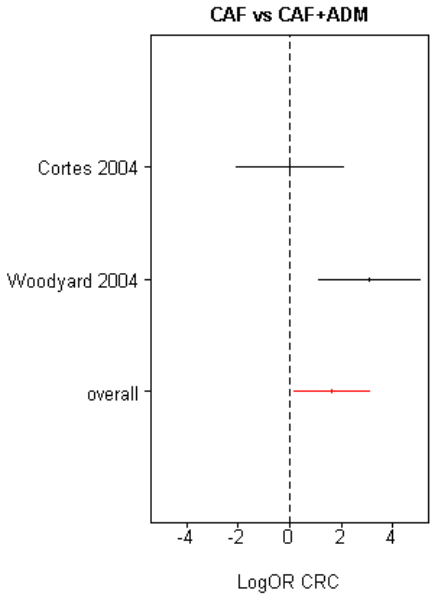


Standard pair-wise Bayesian meta-analysis – CAF vs CAF+ADM		
Study	***LogOR***	***90% CI****
Cortes et al. (2006)	0.00	-2.12, 2.12
Woodyard et al. (2004)	3.09	1.11, 5.07
	LogOR*	***90% CrI*****
Overall (FE)	1.64	0.19, 3.09

*90% CI is the 90% Confidence Interval for the LogOR.

**LogOR= Log Odds Ratio of the posterior distribution of the Bayesian meta-analysis.

***90% CrI is the 90% Credible Interval for LogOR.

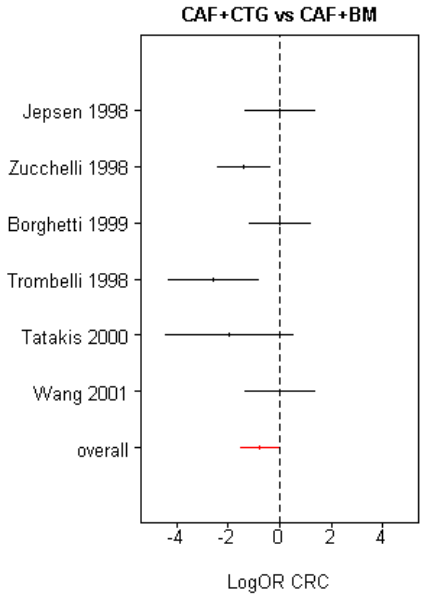


Standard pair-wise Bayesian meta-analysis – CAF+CTG vs CAF+BM		
Study	***LogOR***	***90% CI****
Jepsen et al. (1998)	0.00	-1.35, 1.35
Zucchelli et al. (1998)	-1.39	-2.39, -0.39
Borghetti et al. (1999)	0.00	-1.20, 1.20
Trombelli et al. (1998)	-2.58	-4.35, -0.81
Tatakis & Trombelli (2000)	-1.95	-4.43, 0.53
Wang et al. (2001)	0.00	-1.35, 1.35
	LogOR*	***90% CrI*****
Overall (RE)	-0.78	-1.50, -0.07
Tau^2	0.10	0.00, 2.13
I^2	9.83	0.18, 69.21

*90% CI is the 90% Confidence Interval for the LogOR.

**LogOR is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and I^2 where LogOR is the median.

***90% CrI is the 90% Credible Interval for the LogOR.

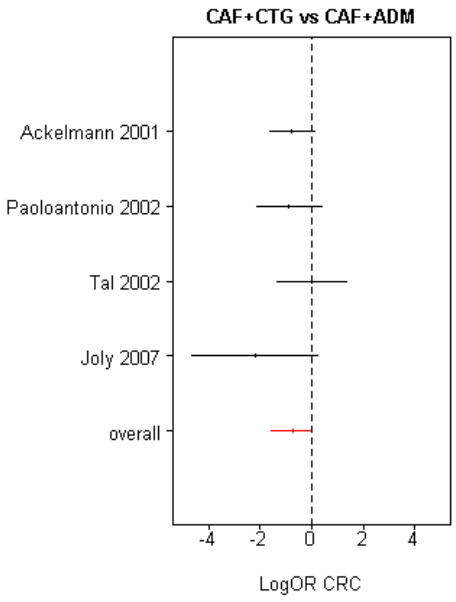


Standard pair-wise Bayesian meta-analysis – CAF+CTG vs CAF+ADM		
Study	***LogOR***	***90% CI****
Aichelmann-Reidy et al. (2001)	-0.76	-1.66, 0.14
Paolantonio et al. (2002)	-0.88	-2.16, 0.40
Tal et al. (2002)	0.00	-1.35, 1.35
Joly et al. (2007)	-2.20	-4.64, 0.24
	LogOR*	***90% CrI*****
Overall (RE)	-0.75	-1.56, 0.02
Tau^2	0.04	0.00, 1.67
I^2	3.55	0.12, 63.70

*90% CI is the 90% Confidence Interval for the LogOR.

**LogOR is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and I^2 where LogOR is the median.

***90% CrI is the 90% Credible Interval for the LogOR.



NM Consistency model - Lu and Ades

Among the total $K(K-1)/2 = 21$ potential pairs of comparisons, $DC = 9$ pairwise comparisons are independently supported by direct evidence from the data (Fig. 25). For describing all possible treatment effects in a model, we need to specify $K-1 = 6$ basic parameters that can form a spanning tree. It would be natural to choose the effects of the six combination treatments (i.e., CAF+CTG, CAF+BM, CAF+EMD, CAF+ADM, CAF+PRP, CAF+HF-DDS) relative to the treatment CAF alone. However, no direct evidence is available for CAF versus CAF+HF-DDS. Treatment CAF+HF-DDS is an isolated vertex (see Fig. 26) compared only with treatment CAF+CTG in trial 22, and thus we must have $d_{CAF+CTG\ CAF+HF\text{-}DDS}$, instead of $d_{CAF\ CAF+HF\text{-}DDS}$, as a basic parameter.

Therefore, we have the following six basic parameters (represented by the solid blue lines in Fig. 26):

$d_{CAF\ CAF+CTG}$

$d_{CAF\ CAF+BM}$

$d_{CAF\ CAF+EMD}$

$d_{CAF\ CAF+ADM}$

$d_{CAF\ CAF+PRP}$

$d_{CAF+CTG\ CAF+HF\text{-}DDS}$

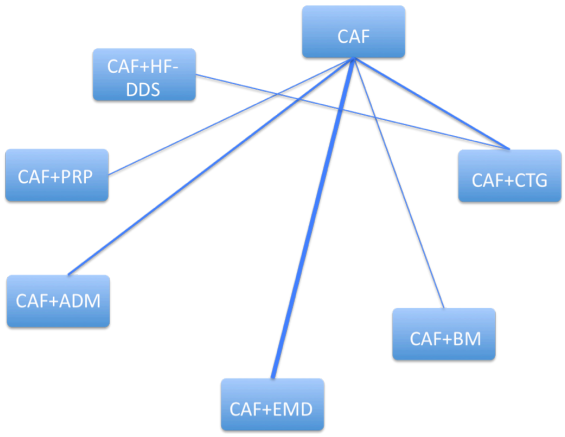


Fig.26 Spanning tree for CRC

Functional Parameters (on the left of each equation below) can be expressed in terms of difference between basic parameters (on the right of each equation below) (Figure 27) as follows:

$$d_{CAF+CTG\ CAF+BM} = d_{CAF\ CAF+CTG} - d_{CAF\ CAF+BM}$$

$$d_{CAF+CTG\ CAF+EMD} = d_{CAF\ CAF+CTG} - d_{CAF\ CAF+EMD}$$

$$d_{CAF+CTG\ CAF+ADM} = d_{CAF\ CAF+CTG} - d_{CAF\ CAF+ADM}$$

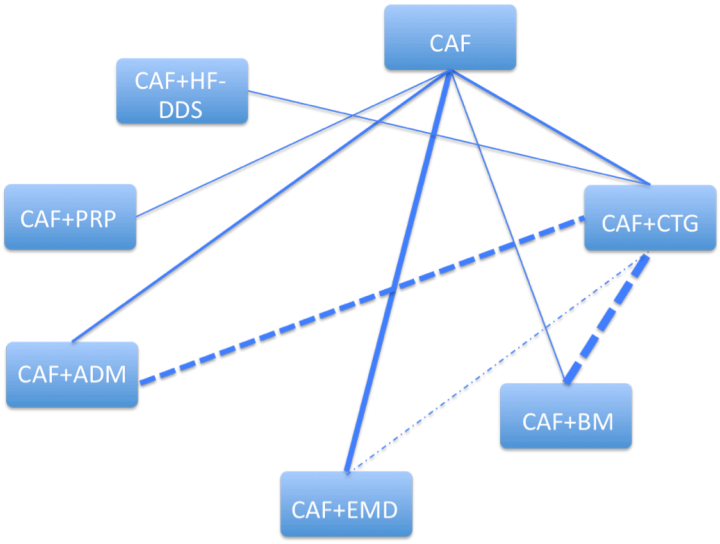


Fig.27 Network of comparisons expressing the relationship between Basic and Functional parameters. The solid lines represent comparisons whose treatment contrasts are specified as basic parameters. All solid lines form a spanning tree. The dotted lines represent comparisons associated with functional parameters.

NM Inconsistency model - Lu and Ades

Inconsistency Degree of Freedom (ICDF)

Potential number of inconsistencies as defined by Lu and Ades (2006) can be calculated as follows in the network diagram for CRC:

ICDF (for 2-arm trials) = #Functional Parameters = DC – K +1

ICDF for CRC = #Functional Parameters = 9 – 7 +1 = 3

Three (3) inconsistencies are identified in the NM for CRC.

Inconsistency Factors (ICFs)/w-factors

Discrepancy between the Functional Parameters and the difference between the Basic Parameters can represented as follows:

$$d_{CAF+CTG\ CAF+BM} = d_{CAF\ CAF+CTG} - d_{CAF\ CAF+BM} + W_{CAF\ CAF+CTG\ CAF+BM}$$

$$d_{CAF+CTG\ CAF+EMD} = d_{CAF\ CAF+CTG} - d_{CAF\ CAF+EMD} + W_{CAF\ CAF+CTG\ CAF+EMD}$$

$$d_{CAF+CTG\ CAF+ADM} = d_{CAF\ CAF+CTG} - d_{CAF\ CAF+ADM} + W_{CAF\ CAF+CTG\ CAF+ADM}$$

Three (3) ICFs (w-factors) are defined and attached to the corresponding evidence cycles (Figure 28):

$W_{CAF\ CAF+CTG\ CAF+BM}$; $W_{CAF\ CAF+CTG\ CAF+EMD}$; $W_{CAF\ CAF+CTG\ CAF+ADM}$

The number of w-factors (3, for CRC analysis) being estimated is not equal to the number of treatment comparisons (9, for CRC analysis) (as showed by Lumley model) informed by data but is defined according to the presence of Functional Parameters.

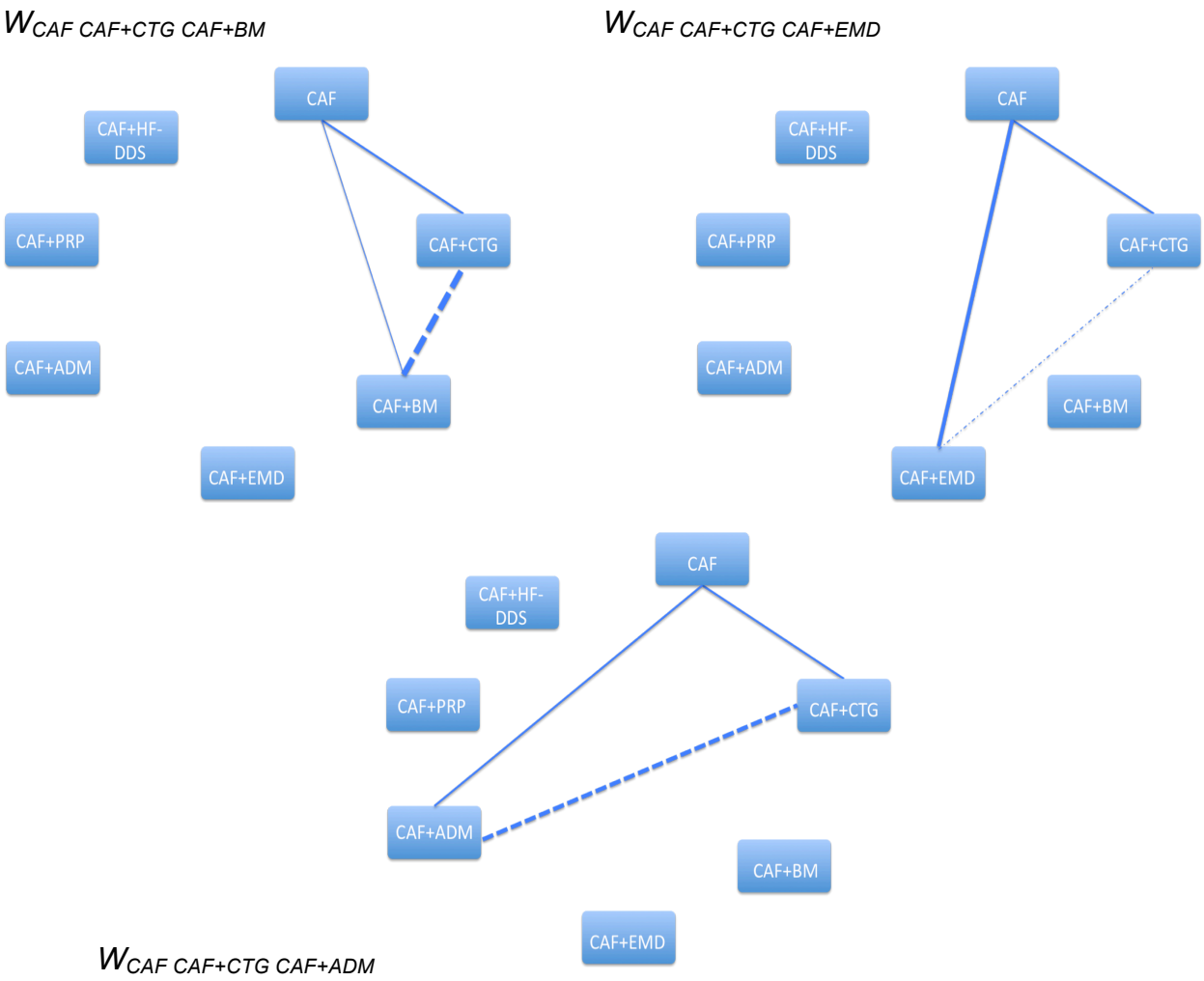


Fig.28 Three (3) evidence cycles are defined by relations between Basic and Functional parameters for the Network Meta-analysis under Consistency model. The solid lines represent comparisons whose treatment contrasts are specified as basic parameters. The dotted lines represent comparisons associated with functional parameters.

Results of NM models with and without Inconsistency Factors

NM random-effects Consistency model show an higher estimate of the median of the posterior distribution of the between-trials variance (Tau^2) then Inconsistency model (6.83 and 0.16 respectively). Results with and without the inconsistency factors are given in Table 18, 20, 22 (Figure 29, 31).

The goodness-of-fit statistic, DIC, for both models is similar (68.97 and 69.99 respectively) even if a little higher in the Inconsistency model.

In Inconsistency model, the values of the overall Inconsistency, $Tau\text{-}w^2$ = 1.07, and Inconsistency probability, $Pr\ (Tau\text{-}w^2 > Tau^2)$ = 0.78, suggest the presence of Inconsistency between sources of evidence on posterior treatment effects.

The values of $W_{CAF\ CAF+CTG\ CAF+BM}$ = 0.35, $W_{CAF\ CAF+CTG\ CAF+EMD}$ = 0.20 and $W_{CAF\ CAF+CTG\ CAF+ADM}$ = - 0.66 suggest the presence of Inconsistency in the three evidence cycles.

Combinations of CAF+EMD and CAF+CTG showed the best results in terms of CRC under both Consistency and Inconsistency models and occuped the first and the second position respectively in the ranking of treatments. (Table 19, 21; Figure 30, 32)

Mean residual deviance for individual data points in models with and without inconsistency factors has been compared and Figure 33 shows the presence of three outliers, corresponding to trials 4, 9 and 14, in the model without w-factors. Point 4 correspond to the comparison CAF vs CAF+EMD of the trial of Modica et al. 2000, point 9 to the comparison CAF vs CAF+ADM of the trial of Woodyard et al. 2004 and point 14 to the comparison CAF+CTG vs CAF+BM of the trial of Trombelli et al. 1998. The last one confirm suspicions about the (CAF CAF+CTG CAF+BM) cycle.

Another Inconsistency model has been carried out after removing points 4 , 9 and 14 from the analysis. The results (Table 23) show both ICFs ($W_{CAF\ CAF+CTG\ CAF+BM}$ = 0.35, $W_{CAF\ CAF+CTG\ CAF+EMD}$ = - 0.02, $W_{CAF\ CAF+CTG\ CAF+ADM}$ = 0.02) and overall Inconsistency ($Tau\text{-}w^2$ = 0.64) are substantially reduced, not the Inconsistency probability due to the simoutaneous reduction of between-trials variance (Tau^2 = 0.09). Results of Consistency and Inconsistency (full data and 4, 9 and 14 deleted) models are showed in Table 24.

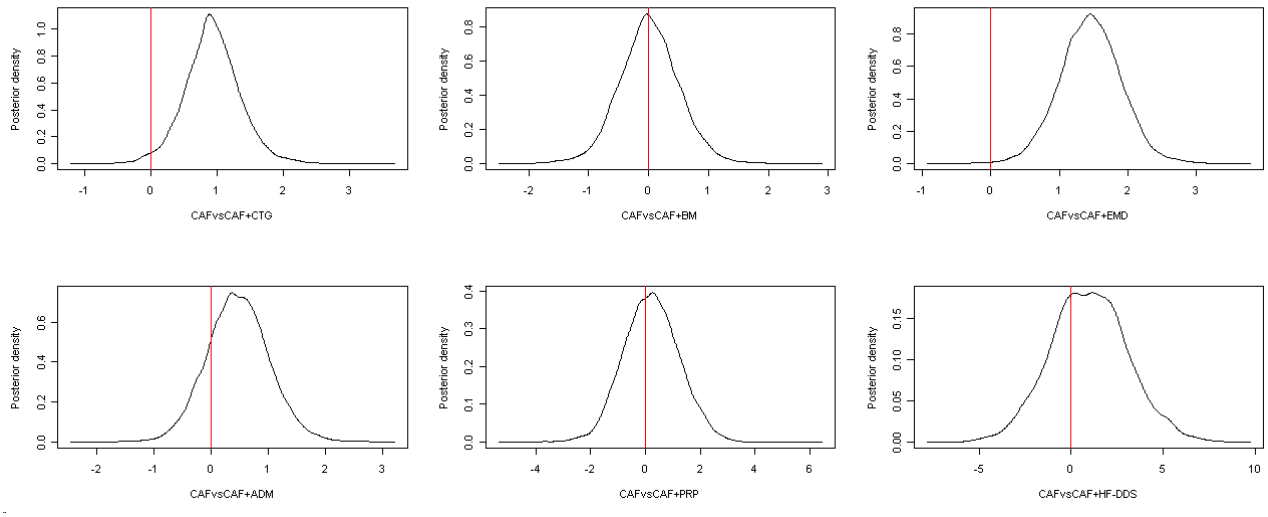
Tab.18 NM for CRC: Consistency model (without ICFs/w-factors). Results of all possible pair-wise treatment comparisons.

Network Meta-analysis – Consistency model		
Treatment Comparison	***LogOR****	***90% CrI****
CAF vs CAF+CTG	0.93	0.26, 1.62
CAF vs CAF+BM	0.01	-0.79, 0.84
CAF vs CAF+EMD	1.45	0.71, 2.18
CAF vs CAF+ADM	0.47	-0.40, 1.37
CAF vs CAF+PRP	0.22	-1.38, 1.91
CAF vs CAF+HF-DDS	0.98	-2.47, 4.48
CAF+CTG vs CAF+BM	-0.92	-1.53, -0.30
CAF+CTG vs CAF+EMD	0.52	-0.40, 1.41
CAF+CTG vs CAF+ADM	-0.46	-1.17, 0.27
CAF+CTG vs CAF+PRP	-0.71	-2.47, 1.09
CAF+CTG vs CAF+HF-DDS	0.05	-3.28, 3.50
CAF+BM vs CAF+EMD	1.44	0.42, 2.45
CAF+BM vs CAF+ADM	0.46	-0.48, 1.39
CAF+BM vs CAF+PRP	0.21	-1.63, 2.08
CAF+BM vs CAF+HF-DDS	0.97	-2.45, 4.47
CAF+EMD vs CAF+ADM	-0.98	-2.04, 0.11
CAF+EMD vs CAF+PRP	-1.23	-3.06, 0.61
CAF+EMD vs CAF+HD-DDS	-0.47	-3.93, 3.10
CAF+ADM vs CAF+PRP	-0.25	-2.11, 1.64
CAF+ADM vs CAF+HF-DDS	0.51	-2.87, 4.03
CAF+PRP vs CAF+HF-DDS	0.76	-3.01, 4.71
Tau^2	6.83	0.99, 572.28
DIC	68.97	

*LogOR is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 where LogOR is the median.

**90% CrI is the 90% Credible Interval for LogOR.

Fig.29 NM for CRC: Plots of the posterior densities of each treatment compared to CAF under Consistency model (without ICFs/w-factors).



Tab.19 NM for CRC: Ranking and Best for the seven treatments included in the analysis under Consistency model (without ICFs/w-factors).

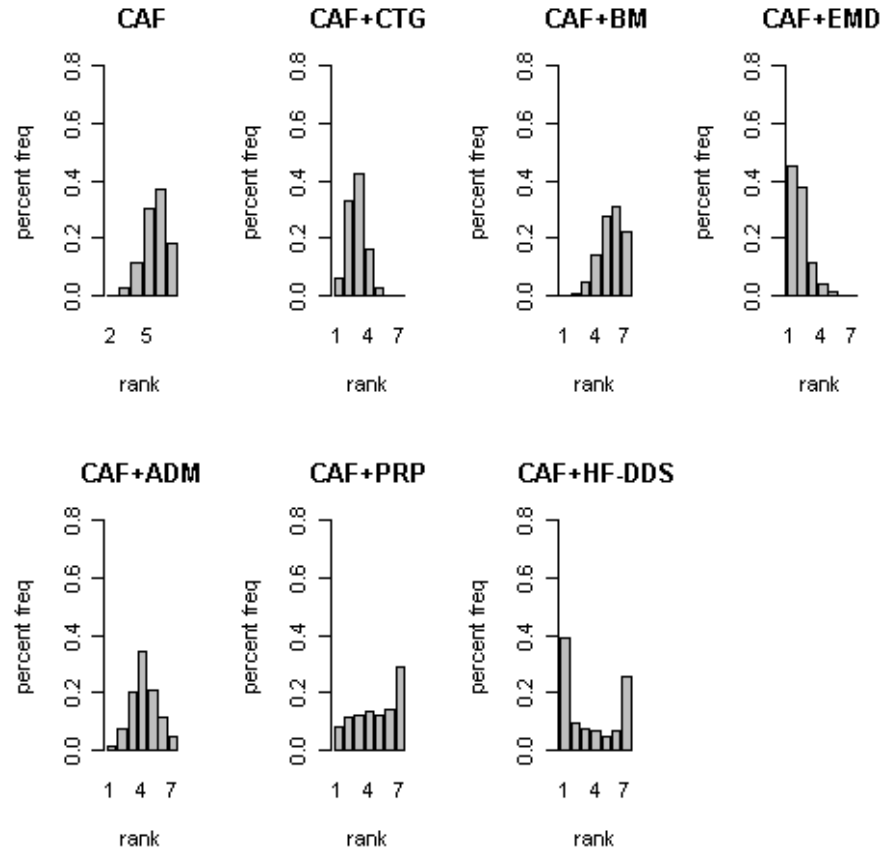
	Ranking		***Best***
Treatment	***LogOR****	***90% CrI****	***Pr******
CAF	5.56	4.00, 7.00	0.00
CAF+CTG	2.77	1.00, 4.00	0.06
CAF+BM	5.51	3.00, 7.00	<0.00
CAF+EMD	1.78	1.00, 4.00	0.46
CAF+ADM	4.19	2.00, 6.00	0.02
CAF+PRP	4.67	1.00, 7.00	0.08
CAF+HF-DDS	3.52	1.00, 7.00	0.39

*LogOR is the median of the posterior distribution of the Bayesian meta-analysis.

**90% CrI is the 90% Credible Interval for the LogOR.

***Pr is the probability that each treatment is the best.

Fig.30 NM for CRC: Ranking for the seven treatments under Consistency model (without ICFs/w-factors).



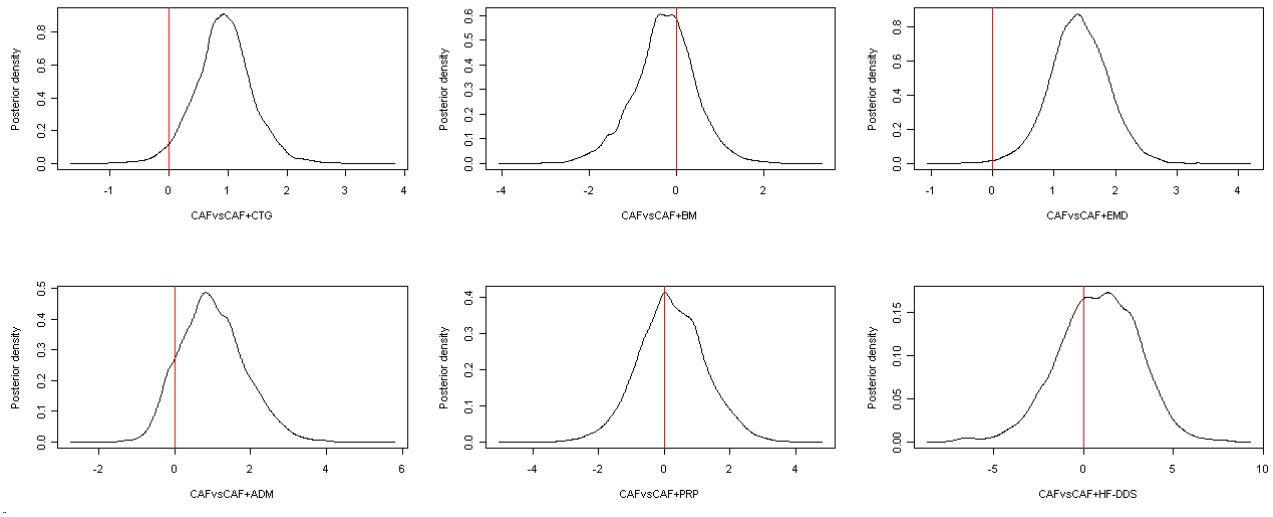
Tab.20 NM for CRC: Inconsistency model (with ICFs/w-factors). Results of all possible pair-wise treatment comparisons.

Network Meta-analysis – Inconsistency model (with ICFs/w-factors)		
Treatment Comparison	***LogOR****	***90% CrI****
CAF vs CAF+CTG	0.93	0.13, 1.73
CAF vs CAF+BM	-0.27	-1.54, 0.87
CAF vs CAF+EMD	1.40	0.62, 2.18
CAF vs CAF+ADM	1.00	-0.28, 2.50
CAF vs CAF+PRP	0.23	-1.43, 1.95
CAF vs CAF+HF-DDS	0.82	-2.87, 4.27
CAF+CTG vs CAF+BM	-1.19	-2.51, -0.06
CAF+CTG vs CAF+EMD	0.48	-0.56, 1.53
CAF+CTG vs CAF+ADM	0.07	-1.22, 1.74
CAF+CTG vs CAF+PRP	-0.70	-2.55, 1.19
CAF+CTG vs CAF+HF-DDS	-0.11	-3.71, 3.23
CAF+BM vs CAF+EMD	1.67	0.31, 3.03
CAF+BM vs CAF+ADM	1.27	-0.35, 3.26
CAF+BM vs CAF+PRP	0.50	-1.57, 2.54
CAF+BM vs CAF+HF-DDS	1.09	-2.70, 4.72
CAF+EMD vs CAF+ADM	-0.40	-1.90, 1.28
CAF+EMD vs CAF+PRP	-1.17	-2.99, 0.64
CAF+EMD vs CAF+HF-DDS	-0.58	-4.44, 2.97
CAF+ADM vs CAF+PRP	-0.77	-2.99, 1.37
CAF+ADM vs CAF+HF-DDS	-0.18	-4.13, 3.52
CAF+PRP vs CAF+HF-DDS	0.59	-3.59, 4.42
w-CAF.CTG.BM	0.35	-0.75, 1.73
w-CAF.CTG.EMD	0.20	-1.07, 1.64
w-CAF.CTG.ADM	-0.66	-2.51, 0.53
Tau^2	0.16	0.00, 1.09
Tau-w^2	1.07	0.01, 26.02
Pr (Tau-w^2 > Tau^2)	0.78	
DIC	69.99	

*LogOR is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and Tau-w^2 where LogOR is the median.

**90% CrI is the 90% Credible Interval for the LogOR.

Fig.31 NM for CRC: Plots of posterior densities of each treatment compared to CAF under the Inconsistency model (with ICFs/w-factors).



Tab.21 NM for CRC: Ranking and Best for the seven treatments included in the analysis under Inconsistency model (with ICFs/w-factors).

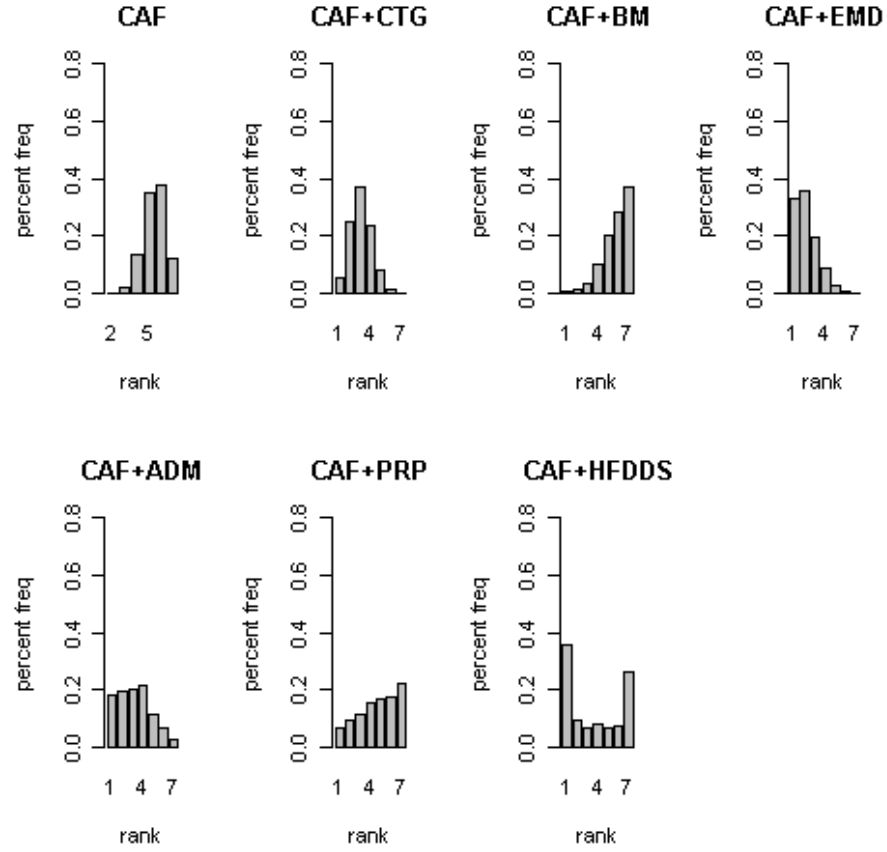
	Ranking		***Best***
Treatment	***LogOR****	***90% CrI****	***Pr*****
CAF	5.44	4.00, 7.00	0.00
CAF+CTG	3.09	1.00, 5.00	0.06
CAF+BM	5.81	4.00, 7.00	0.00
CAF+EMD	2.12	1.00, 4.00	0.33
CAF+ADM	3.20	1.00, 6.00	0.18
CAF+PRP	4.68	1.00, 7.00	0.07
CAF+HF-DDS	3.68	1.00, 7.00	0.36

*LogOR is the median of the posterior distribution of the Bayesian meta-analysis.

**90% CrI is the 90% Credible Interval for the LogOR.

***Pr is the probability that each treatment is the best.

Fig. 32 NM for CRC: Ranking for the seven treatments under Inconsistency model (with ICFs/w-factors).



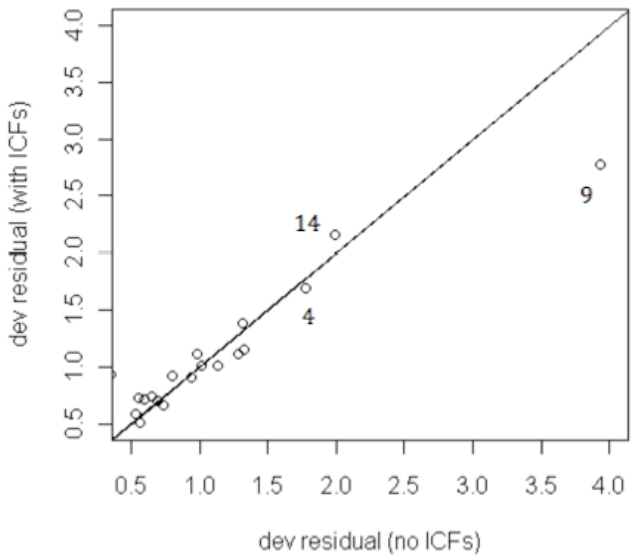
Tab.22 NM for CRC: Standard meta-analysis, Consistency and Inconsistency models (limited to the comparisons informed by data and analyzed in SM).

	SM	***NM (Consistency model)***	***NM (Inconsistency model)***
Treatment Comparison (N*)	***LogOR (90% CrI)*****	***LogOR (90% CrI)***	***LogOR (90% CrI)***
CAF vs CAF+CTG (2)	0.91 (0.23, 1.61)	0.93 (0.26, 1.62)	0.93 (0.13, 1.73)
CAF vs CAF+EMD (4)	1.36 (0.43, 2.25)	1.45 (0.71, 2.18)	1.40 (0.62, 2.18)
CAF vs CAF+ADM (2)	1.64 (0.19, 3.09)	0.47 (-0.40, 1.37)	1.00 (-0.28, 2.50)
CAF+CTG vs CAF+BM (6)	-0.78 (-1.50, -0.07)	-0.92 (-1.53, -0.30)	-1.19 (-2.51, -0.06)
CAF+CTG vs CAF+ADM (4)	-0.75 (-1.56, 0.02)	-0.46 -1.17, 0.27	0.07 (-1.22, 1.74)
Tau2	-	6.83 (0.99, 572.28)	0.16 (0.00, 1.09)
Tau-w2	-	-	1.07 (0.01, 26.02)
Pr (Tau-w2 > Tau2)	-	-	0.78
DIC	-	68.97	69.99

*N is the number of trials with direct evidence for each treatment comparison.

** LogOR is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and $Tau\text{-}w^2$ where LogOR is the median.

Fig.33 NM for CRC: residuals in fitting Consistency models without w-factors and NM Inconsistency model (with ICFs). Points 4, 9 and 14 correspond to the comparison CAF vs CAF+EMD of the trial of Modica et al. (2000), CAF vs CAF+ADM of the trial of Woodyard et al. (2004) and CAF+CTG vs CAF+BM of the trial of Trombelli et al. (1998) respectively.



Tab.23 NM for CRC: Inconsistency model (4, 9 and 14 deleted). Results of all possible pair-wise treatment comparisons.

Network Meta-analysis – Inconsistency model (4, 9 and 14 deleted)		
Treatment Comparison	***LogOR****	***90% CrI****
CAF vs CAF+CTG	0.76	0.01, 1.52
CAF vs CAF+BM	-0.23	-1.32, 0.79
CAF vs CAF+EMD	1.66	0.85, 2.48
CAF vs CAF+ADM	0.01	-1.48, 1.52
CAF vs CAF+PRP	0.26	-1.38, 1.80
CAF vs CAF+HF-DDS	0.66	-2.77, 4.04
CAF+CTG vs CAF+BM	-1.00	-2.27, 0.06
CAF+CTG vs CAF+EMD	0.90	-0.11, 1.92
CAF+CTG vs CAF+ADM	-0.75	-2.26, 0.73
CAF+CTG vs CAF+PRP	-0.51	-2.32, 1.18
CAF+CTG vs CAF+HF-DDS	-0.11	-3.40, 3.16
CAF+BM vs CAF+EMD	1.90	0.62, 3.26
CAF+BM vs CAF+ADM	0.24	-1.39, 2.12
CAF+BM vs CAF+PRP	0.49	-1.42, 2.37
CAF+BM vs CAF+HF-DDS	0.89	-2.65, 4.35
CAF+EMD vs CAF+ADM	-1.65	-3.31, 0.04
CAF+EMD vs CAF+PRP	-1.41	-3.22, 0.30
CAF+EMD vs CAF+HF-DDS	-1.01	-4.53, 2.48
CAF+ADM vs CAF+PRP	0.25	-2.03, 2.43
CAF+ADM vs CAF+HF-DDS	0.65	-3.09, 4.30
CAF+PRP vs CAF+HF-DDS	0.40	-3.45, 4.18
w-CAF.CTG.BM	0.35	-0.65, 1.69
w-CAF.CTG.EMD	-0.02	-1.25, 1.23
w-CAF.CTG.ADM	0.02	-1.43, 1.45
Tau^2	0.09	0.00, 0.75
Tau-w^2	0.64	0.00, 19.68
Pr (Tau-w^2 > Tau^2)	0.77	
DIC	54.50	

*LogOR is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and $Tau\text{-}w^2$ where LogOR is the median.

**90% CrI is the 90% Credible Interval for the LogOR.

Tab.24 NM for CRC: Consistency, Inconsistency (full data) and Inconsistency (4, 9 and 14 deleted) models (limited to the comparisons informed by data and analyzed in SM).

		NM (Inconsistency model)	
	NM (Consistency model)	***Full data***	***4, 9 and 14 deleted***
Treatment Comparison (N*)	***LogOR (90% CrI)****	***LogOR (90% CrI)***	***LogOR (90% CrI)***
CAF vs CAF+CTG (2)	0.93 (0.26, 1.62)	0.93 (0.13, 1.73)	0.76 (0.01, 1.52)
CAF vs CAF+EMD (4)***	1.45 (0.71, 2.18)	1.40 (0.62, 2.18)	1.66 (0.85, 2.48)
CAF vs CAF+ADM (2)****	0.47 (-0.40, 1.37)	1.00 (-0.28, 2.50)	0.01 (-1.48, 1.52)
CAF+CTG vs CAF+BM (6)*****	-0.92 (-1.53, -0.30)	-1.19 (-2.51, -0.06)	-1.00 (-2.27, 0.06)
CAF+CTG vs CAF+ADM (4)	-0.46 -1.17, 0.27	0.07 (-1.22, 1.74)	-0.75 (-2.26, 0.73)
w-CAF.CTG.BM	-	0.35 (-0.75, 1.73)	0.35 (-0.65, 1.69)
w-CAF.CTG.EMD	-	0.20 (-1.07, 1.64)	-0.02 (-1.25, 1.23)
w-CAF.CTG.ADM	-	-0.66 (-2.51, 0.53)	0.02 (-1.43, 1.45)
Tau^2	6.83 (0.99, 572.28)	0.16 (0.00, 1.09)	0.09 (0.00, 0.75)
$Tau\text{-}w^2$	-	1.07 (0.01, 26.02)	0.64 (0.00, 19.68)
Pr ($Tau\text{-}w^2$ > Tau^2)	-	0.78	0.77
DIC	68.97	69.99	54.50

*N is the number of trials with direct evidence for each treatment comparison.

**LogOR is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and $Tau\text{-}w^2$ where LogOR is the median (90% CrI is the 90% Credible Interval).

***In the Inconsistency model without 4, 9 and 14, N = 3.

****In the Inconsistency model without 4, 9 and 14, N = 1.

*****In the Inconsistency model without 4, 9 and 14, N = 5.

NM Inconsistency model - Bayesian extension of Lumley's method

Among the total $K(K-1)/2 = 21$ potential pairs of comparisons, $DC = 9$ pairwise comparisons are independently supported by direct evidence from the data (Fig. 25).

The number of terms w_{kb} being estimated is equal to the number of treatment comparisons. Thus, nine (9) w-factors can identified for CRC analysis:

$w_{CAF\ CAF+CTG}$; $w_{CAF\ CAF+BM}$; $w_{CAF\ CAF+EMD}$; $w_{CAF\ CAF+ADM}$; $w_{CAF\ CAF+PRP}$;

$w_{CAF+CTG\ CAF+BM}$; $w_{CAF+CTG\ CAF+EMD}$; $w_{CAF+CTG\ CAF+ADM}$; $w_{CAF+CTG\ CAF+HF\text{-}DDS}$

Results for the Bayesian extension of Lumley's method

The median of the posterior distribution of the between-trials variance (Tau^2) is very small (0.15) and similar to Inconsistency model previously described. Results for all possible pair-wise comparisons are given in Table 25. Posterior densities of each treatment compared to reference treatment (CAF) are represented in Figure 34.

The goodness-of-fit statistic, DIC, for this NM model (70.19) is close to the value of Inconsistency model by Lu and Ades (69.99).

The lower values (with respect to Lu and Ades model) of the overall Inconsistency, $Tau\text{-}w^2 = 0.46$, and Inconsistency probability, $Pr\ (Tau\text{-}w^2 > Tau^2) = 0.69$, may be explained according to the fact that in the present model the w-factors are attached to single comparisons informed by data and not to evidence cycles.

The values of w-factors relative to CAFvsCAF+BM (- 0.16), CAFvsCAF+ADM (0.24), CAF+CTGvsCAF+BM (0.18), CAF+CTGvsCAF+ADM (- 0.46) and asimmetries in the relative posterior densities (Figure 36) suggest the presence of Inconsistency in correspondence of these comparisons (Table 25).

Combinations of CAF+EMD and CAF+CTG showed the best results in terms of CRC under the Bayesian extension of Lumley's method too (Table 26, Figure 35).

Mean residual deviance (Figure 37) for individual data points in the Bayesian extension of Lumley's method too show similar results to the Inconsistency model by Lu and Ades.

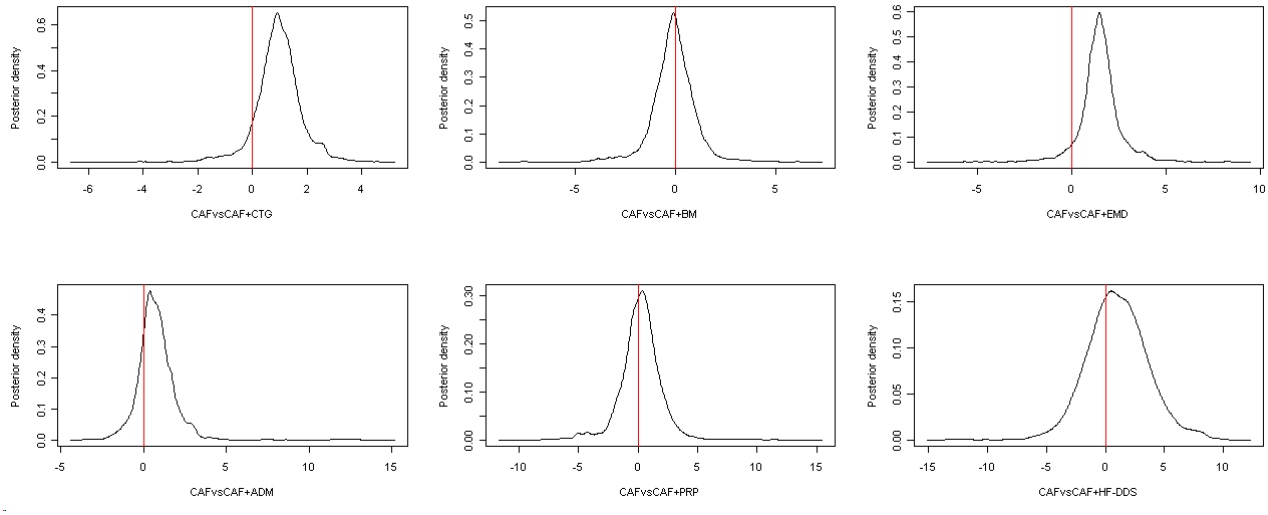
Tab.25 NM for CRC: Bayesian extension of Lumley's method. Results of all possible pair-wise treatment comparisons.

Network Meta-analysis – Bayesian extension of Lumley's method				
Treatment Comparison	***LogOR****	***90% CrI****	***w***	***90% CrI***
CAF vs CAF+CTG	0.94	-0.38, 2.34	-0.02	-1.42, 1.31
CAF vs CAF+BM	-0.09	-1.78, 1.52	-0.16	-1.75, 1.33
CAF vs CAF+EMD	1.50	-0.15, 3.28	-0.08	-1.76, 1.43
CAF vs CAF+ADM	0.81	-0.88, 2.68	0.24	-1.05, 2.08
CAF vs CAF+PRP	0.24	-2.50, 2.84	-0.01	-1.98, 2.13
CAF vs CAF+HF-DDS	1.00	-3.04, 5.35	-	-
CAF+CTG vs CAF+BM	-1.04	-2.82, 0.49	0.18	-1.26, 1.99
CAF+CTG vs CAF+EMD	0.56	-1.19, 2.36	0.08	-1.48, 1.62
CAF+CTG vs CAF+ADM	-0.14	-1.68, 1.62	-0.46	-2.35, 0.90
CAF+CTG vs CAF+PRP	-0.71	-3.68, 2.16	-	-
CAF+CTG vs CAF+HF-DDS	0.06	-3.88, 4.22	-0.11	-2.37, 1.65
CAF+BM vs CAF+EMD	1.59	-0.50, 4.12	-	-
CAF+BM vs CAF+ADM	0.90	-0.99, 3.37	-	-
CAF+BM vs CAF+PRP	0.33	-2.83, 3.48	-	-
CAF+BM vs CAF+HF-DDS	1.10	-3.04, 5.46	-	-
CAF+EMD vs CAF+ADM	-0.70	-2.86, 1.73	-	-
CAF+EMD vs CAF+PRP	-1.27	-4.61, 1.81	-	-
CAF+EMD vs CAF+HF-DDS	-0.50	-4.71, 4.01	-	-
CAF+ADM vs CAF+PRP	-0.57	-3.86, 2.26	-	-
CAF+ADM vs CAF+HF-DDS	0.20	-4.02, 4.59	-	-
CAF+PRP vs CAF+HF-DDS	0.77	-4.23, 6.21	-	-
Tau^2	0.15	0.00, 1.06	-	-
Tau-w^2	0.46	0.01, 8.59		
Pr (Tau-w^2 > Tau^2)	0.69			
DIC	70.19			

*LogOR is the mean of the posterior distribution of the Bayesian meta-analysis except for Tau^2 and Tau-w^2 where LogOR is the median.

**90% CrI is the 90% Credible Interval for the LogOR.

Fig.34 NM for CRC: Plots of posterior densities of each treatment compared to CAF under the Bayesian extension of Lumley's method.



Tab.26 NM for CRC: Ranking and Best for the seven treatments included in the analysis under the Bayesian extension of Lumley's method.

	Ranking		***Best***
Treatment	***LogOR****	***90% CrI****	***Pr*****
CAF	5.38	3.00, 7.00	0.00
CAF+CTG	3.09	1.00, 5.00	0.06
CAF+BM	5.39	3.00, 7.00	0.01
CAF+EMD	2.26	1.00, 5.00	0.38
CAF+ADM	3.76	1.00, 6.00	0.09
CAF+PRP	4.55	1.00, 7.00	0.11
CAF+HF-DDS	3.62	1.00, 7.00	0.35

*LogOR is the median of the posterior distribution of the Bayesian meta-analysis.

**90% CrI is the 90% Credible Interval for the LogOR.

***Pr is the probability that each treatment is the best.

Fig.35 NM for CRC: Ranking for the seven treatments under Bayesian extension of Lumley's method.

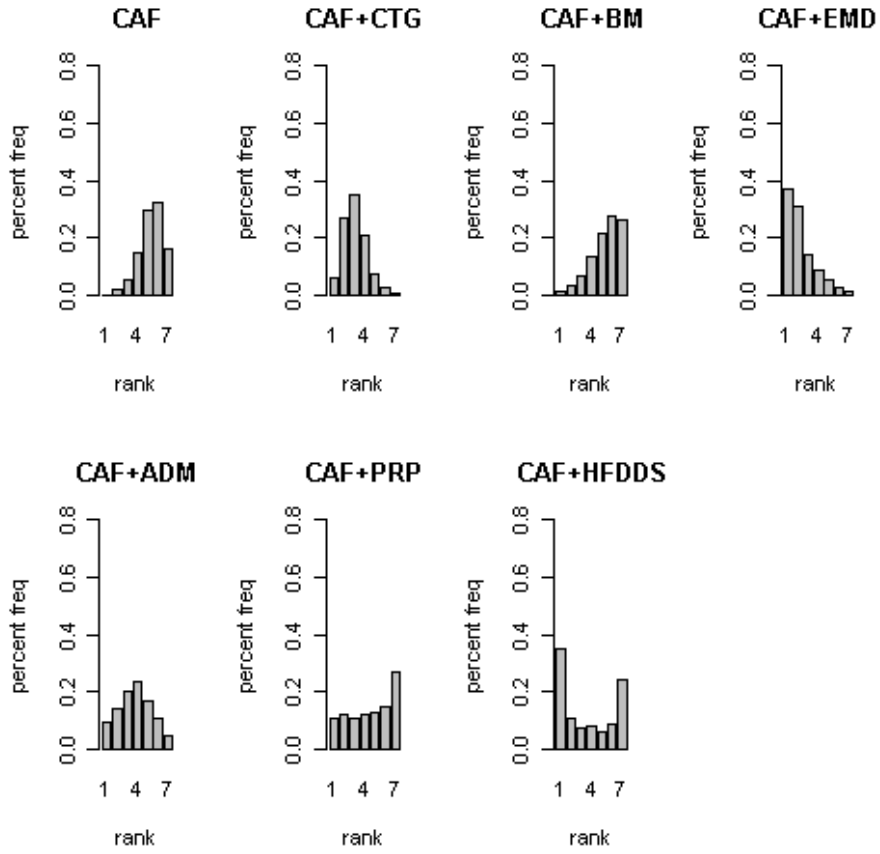


Fig.36 NM for CRC: Plots of posterior densities of w-factors under the Bayesian extension of Lumley's method.

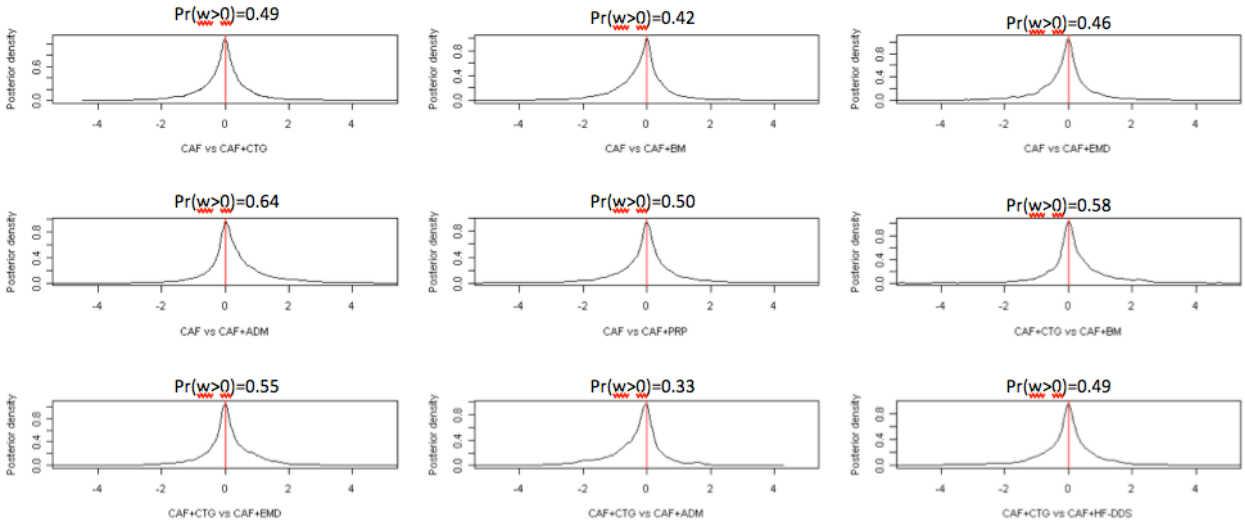
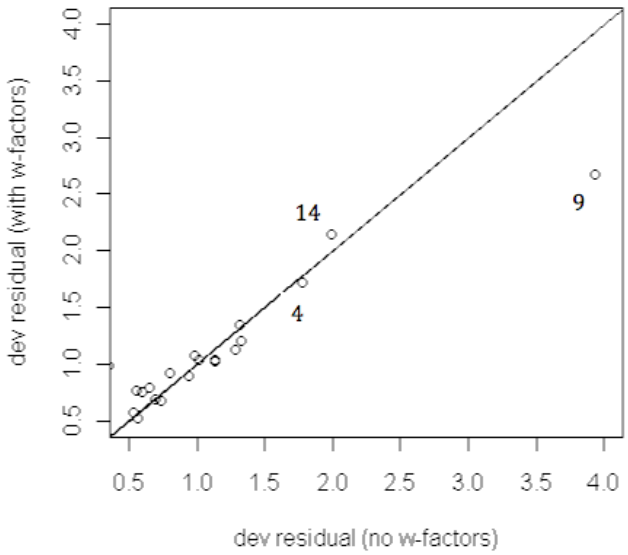


Fig.37 NM for CRC: residuals in fitting Consistency models without w-factors and NM Bayesian extension of Lumley's method (with w-factors). Points 4, 9 and 14 correspond to the comparison CAF vs CAF+EMD of the trial of Modica et al. (2000), CAF vs CAF+ADM of the trial of Woodyard et al. (2004) and CAF+CTG vs CAF+BM of the trial of Trombelli et al. (1998) respectively.



Kullback Leibler distance

Kullback-Leibler distance between the posterior distribution of treatment effect obtained by SM model (assumed as reference distribution) and by NM Consistency model was calculated for each treatment comparison informed by data and analyzed in single Bayesian meta-analyses (Table 27). KL distance (and relative *q*) for CAFvsCAF+ADM (1.06;0.97) and CAF+CTGvsCAF+ADM (0.18;0.78) and corresponding plots (Figure 38) suggest the presence of more pronunced discrepancy between direct and indirect evidence for these comparisons according to the results of the Bayesian extension of Lumley's method. KL distances (*q*) and w-factors (as derived by the Bayesian extension of Lumley's method) are compared in Table 28.

Tab.27 NM for CRC: Consistency model (without ICFs/w-factors) and standard pair-wise Bayesian meta-analysis (limited to the comparisons informed by data and analyzed in SM).

	Network Meta-analysis (Consistency model)		***Standard pair-wise meta-analysis (Ref.)***		
Treatment Comparison (N*)	***LogOR*****	***90% CrI******	***LogOR***	***90% CrI***	***KL(q)*******
CAF vs CAF+CTG (2)	0.93	0.26, 1.62	0.91	0.23, 1.61	0.00 (0.52)
CAF vs CAF+EMD (4)	1.45	0.71, 2.18	1.36	0.43, 2.25	0.11 (0.72)
CAF vs CAF+ADM (2)	0.47	-0.40, 1.37	1.64	0.19, 3.09	1.06 (0.97)
CAF+CTG vs CAF+BM (6)	-0.92	-1.53, -0.30	-0.78	-1.50, -0.07	0.08 (0.69)
CAF+CTG vs CAF+ADM (4)	-0.46	-1.17, 0.27	-0.75	-1.56, 0.02	0.18 (0.78)

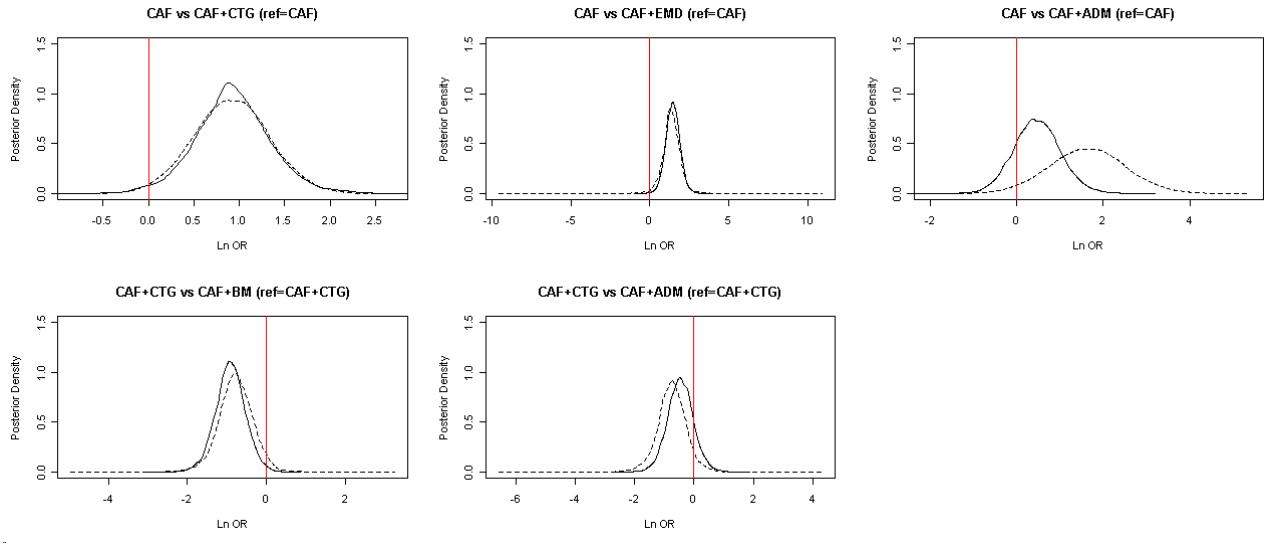
*N is the number of trials with direct evidence for the treatment comparison.

**LogOR is the mean of the posterior distribution of the Bayesian meta-analysis.

***90% CrI is the 90% Credible Interval for the LogOR.

****KI is the Kullback–Leibler distance between the posterior distribution of the Mean for the pair-wise comparison under the NM model and the correspondent pair-wise comparison under the standard meta-analysis model; q is the calibrated value of the Kullback–Leibler distances espressed as a probability.

Fig.38 NM for CRC: Plots of posterior distributions of Consistency model (without ICFs/w-factors) and standard pair-wise Bayesian meta-analysis (limited to the comparisons informed by data and analyzed in SM).



Dotted lines = Network Meta-analysis; solid lines = Standard pair-wise Bayesian Meta-analysis.

Tab.28 NM for CRC: SM, NM Consistency (Lu and Ades) and NM (Bayesian extension of Lumley's method) models (limited to the comparisons informed by data and analyzed in SM).

	SM (Ref.)	***NM (Consistency model)***		***NM (Lumley)***	
Treatment Comparison (N*)	***LogOR (90% CrI)*****	***LogOR (90% CrI)***	***KL(q)******	***LogOR (90% CrI)***	***w***
CAF vs CAF+CTG (2)	0.91 (0.23, 1.61)	0.93 (0.26, 1.62)	0.00 (0.52)	0.94 (-0.38, 2.34)	-0.02 (-1.42, 1.31)
CAF vs CAF+EMD (4)	1.36 (0.43, 2.25)	1.45 (0.71, 2.18)	0.11 (0.72)	1.50 (-0.15, 3.28)	-0.08 (-1.76, 1.43)
CAF vs CAF+ADM (2)	1.64 (0.19, 3.09)	0.47 (-0.40, 1.37)	1.06 (0.97)	0.81 (-0.88, 2.68)	0.24 (-1.05, 2.08)
CAF+CTG vs CAF+BM (6)	-0.78 (-1.50, -0.07)	-0.92 (-1.53, -0.30)	0.08 (0.69)	-1.04 (-2.82, 0.49)	0.18 (-1.26, 1.99)
CAF+CTG vs CAF+ADM (4)	-0.75 (-1.56, 0.02)	-0.46 -1.17, 0.27	0.18 (0.78)	-0.14 (-1.68, 1.62)	-0.46 (-2.35, 0.90)
Tau^2	-	6.83 (0.99, 572.28)	-	0.15 (0.00, 1.06)	-
Tau-w^2	-	-	-	0.46 (0.01, 8.59	-
Pr (Tau-w^2 > Tau^2)	-	-	-	0.69	-
DIC	-	68.97	-	70.19	-

*N is the number of trials with direct evidence for each treatment comparison.

**LogOR is the mean of the posterior distribution of the Bayesian meta-analysis (90% CrI is the 90% Credible Interval).

***KI is the Kullback– Leibler distance between the posterior distribution of the Mean for the pair-wise comparison under the NM model and the correspondent pair-wise comparison under the standard meta-analysis model; q is the calibrated value of the Kullback–Leibler distances espressed as a probability.

4 Concluding remarks

4.1 Clinical perspective

In this thesis, different Bayesian Network Meta-analyses models were carried out to combine published information in the context of periodontal treatments of gingival recession. The aim was to compare the effect of CAF alone with the effects of CAF in combination with grafts or specific biomaterials (CTG, BM, EMD, ADM, PRP, HF-DDS).

In contrast to standard pair-wise meta-analysis, NM is a statistical procedure which offers some important advantages in the synthesis of research findings. In presence of several treatments choices for the same clinical condition, which is a very common scenario in periodontal and dental care, NM represent a reliable way to summarize direct and indirect evidence at the same time in a unique model.

NM provided results in agreement with previous frequentist pair-wise meta-analyses by Cairo et al. (2008). However, it permitted to consider a wider evidence base than standard pair-wise meta-analysis drawing conclusions on treatments never directly compared in RCTs without breaking randomization (Lumley 2002, Caldwell 2005, Glenny 2005, Sutton 2008). In particular, in our case study, the two combinations of treatments, CAF+EMD and CAF+CTG, which showed the best results in terms of RecRed and CRC, had never been head-to-head tested in single trials.

Combining the data from a set of different studies, NM approach shares al difficulties with standard meta-analysis. Precise definition of treatment procedures, differences in the chracteristics of the partecipants, duration of the follow-up period, outcome measures, quality assessment criteria and others must be accurately considered and, probably, NM should be used and interpreted with caution. In fact, all these factors are potential sources of heterogeneity among studies and Inconsistency.

In the application presented in this work, heterogeneity and Inconsistency did not show particulary high values in their estimates and, in any case, sources of inconsistencies could be located to specific evidence cycles or edges. Moreover, the most important contribute to Inconsistency was determined by a few number of studies (2 for RecRed analysis and 3 for CRC) identifyied in the sensitività analysis. This seems to assess for reliability of the results.

4.2 Statistical perspective

In this thesis different approaches to NM were evaluated according to recent developments of the literature.

We used two of the most important frameworks dealing with the issue of Inconsistency in NM methods, which are the one proposed by Lu and Ades (2004, 2006) and a Bayesian extension of the method described by Lumley (2002). Lack of applications of such NM models is recognised in the body of literature and the concept of Inconsistency is rarely considered.

Attention focused on investigating the sources of variability commonly identified when combining evidence from a set of several trials with multiple treatment comparisons. One of the main issue of NM models proposed in this work was exploring Inconsistency, which emerges when uncertainty due to discrepancy between direct and indirect inference on pair-wise comparisons is present.

Lu and Ades (2006) define Inconsistency as a property of evidence cycles, not of individual data points (i.e. treatments). A difficult task is that several cycles share common edges. Moreover, for most datasets, the degrees of freedom for assessing Inconsistency (ICDF) is small and measures of σ_w^2 will be highly uncertain.

Lumley (2002) account for Inconsistency in a different way: the number of terms w_{kb} (inconsistencies) being estimated appears to be equal to the number of treatment comparisons rather than the ICDF.

Kullback-Leibler distance may be viewed in the same spirit as the work of Lumley, investigating a way to compare direct inference from standard pair-wise meta-analysis and indirect inference from NM model.

Attention and further research should be focused on establish a unique assessment of single specific inconsistencies (ICFs/w-factors) as properties of cycles or edges and then analyze and eventually remove these sources of uncertainty combining sensitivity analyses and clinical knowledge.

Other approaches, including graphical representations using dissimilarity matrices (Chung et al. 2008), have been developed to assess consistency in NM. Salanti et al. (2008) and Dias et al. (2010) have discussed the strengths and weaknesses of these approaches. Further research is needed, to improve understanding of these methods and encourage their use. As Salanti et al. (2008) point out, a measure for Inconsistency analogous to I_2 would be a welcome addition.

Appendices

We show here the WinBUGS codes for RecRed outcome. Data for *N* = 22 studies and *K* = 6 treatment alternatives are listed as shown in Figure A1 where *t* is the treatment, *b* indicates which treatment is the effective trial-specific 'baseline' treatment in that study, *diff* is the mean treatment difference expressed in millimitres, *var* is the variability (standard error), *arm* is the trial-specific number of arms (only two-arm studies are included in this analysis), *comp* is the comparison and *study* is the reference.

```
t       b       diff    var     arm     comp    study
2       1       0.44    0.27    2       1       "Da Silva 2004"
2       1       0.52    0.23    2       1       "Cortellini 2008"
3       1       -0.2    0.21    2       2       "Amarante 2000"
3       1       -0.4    0.29    2       2       "Lins 2003"
4       1       0.9     0.43    2       3       "Modica 2000"
4       1       0.07    0.25    2       3       "Del Pizzo 2005"
4       1       0.377   0.23    2       3       "Spahr 2005"
4       1       0.91    0.44    2       3       "Castellanos 2006"
4       1       0.93    0.24    2       3       "Pilloni 2006"
5       1       0.08    0.16    2       4       "Cortes 2004"
5       1       1.23    0.38    2       4       "Woodyard 2004"
6       1       -0.2    0.35    2       5       "Huang 2005"
3       2       -0.01   0.23    2       6       "Jepsen 1998"
3       2       -0.9    0.26    2       6       "Trombelli 1998"
3       2       -0.6    0.21    2       6       "Zucchelli 1998"
3       2       0.0     0.37    2       6       "Borghetti 1999"
3       2       -0.4    0.24    2       6       "Tatakis 2000"
3       2       -0.2    0.27    2       6       "Wang 2001"
5       2       -0.5    0.29    2       7       "Ackelmann 2001"
5       2       -0.2    0.35    2       7       "Paoloantonio 2002"
5       2       0.29    0.22    2       7       "Tal 2002"
5       2       -1.4    0.41    2       7       "Joly 2007"
```

Fig. A1. Sample WinBUGS data listing for RecRed outcome.

Appendix A: WinBUGS code for the Bayesian meta-analysis

a

```
# Fixed-effects model for standard Bayesian meta-analysis;

{
   for (i in 1:N) {
      varinv[i] <- 1/(se.hat[i]*se.hat[i]);
      beta.hat[i]  ~    dnorm(b,varinv[i]); # model
          }

   # vague prior for parameter b

   b ~ dnorm(0.0,1.0E-6);


  }
```

b

```
# Random-effects model for standard Bayesian meta-analysis;

{
   for (i in 1:N) {
      varinv[i] <- 1/(se.hat[i]*se.hat[i]);
      beta.hat[i]  ~    dnorm(beta[i],varinv[i]); # model
          beta[i]  ~    dnorm(b,prec)}

   # vague priors for parameter b and for RE standard deviation

   b ~ dnorm(0.0,1.0E-6);
   prec ~ dgamma(0.001,0.001);
   tau2 <- 1/prec;

  }
```

Fig. A2. WinBUGS code for the (a) fixed and (b) random effect standard Bayesian meta-analysis models.

Appendix B: WinBUGS code for the NM models

The a,b,c models in Figure A3 assumes that the degree of between-trials variation in random effect models is the same for all the pair-wise comparisons (Homogeneous variance models). This assumption can be relaxed according to Lu and Ades (2004) as showed in the model d (Heterogenous variance model) to obtain estimates of 15 different variances for each treatment comparison.

a

```
# Random-effects NM Consistency model (Lu and Ades 2006)

model{
        for (i in 1:N){
           prec[i]<- 1/var[i]
           diff[i]~dnorm(delta[i],prec[i])
           delta[i] ~ dnorm(md[i],taud[i])  # trial-specific means diff distributions
           taud[i] <- tau                   # precisions of diff distributions
           md[i] <- d[t[i]] - d[b[i]]}      # means of diff distributions

        # vague priors for basic parameters

        d[1]<-0
        for (k in 2:NT) {d[k] ~ dnorm(0,.00001) }

        # vague prior for RE standard deviation

        sd~dunif(0,10)
        tau<-1/pow(sd,2)
        tau2<-1/tau


        # All pairwise differences

        for (c in 1:NT-1){for (k in (c+1):NT){pairwise[c,k]<-d[k]-d[c]}}

                        }
```

b

```
# Random-effects NM Inconsistency model (Lu and Ades 2006)

model{
        for (i in 1:N){
           prec[i]<- 1/var[i]
           diff[i]~dnorm(delta[i],prec[i])
           delta[i] ~ dnorm(md[i],taud[i])     # trial-specific means diff distributions
           taud[i] <- tau                      # precisions of diff distributions
           md[i] <- d[t[i]] - d[b[i]] + w[i]}  # means of diff distributions

        # vague priors for basic parameters

        d[1]<-0
        for (k in 2:NT) {d[k] ~ dnorm(0,.00001) }

        # vague prior for RE standard deviation

        sd~dunif(0,10)
        tau<-1/pow(sd,2)
        tau2<-1/tau

        # All pairwise differences

        for (i in 1:12){w[i]<-0}
        for (i in 13:18){w[i]<-w123}
        for (i in 19:N){w[i]<-w125}

        for (c in 1:NT-1){for (k in (c+1):NT){pairwise[c,k]<-d[k]-d[c]}}

        # vague prior for w

        w123~dnorm(0,tauw)
        w125~dnorm(0,tauw)
        sdw~dunif(0,10)
        tauw<-1/pow(sdw,2)
        tau2w<-1/tauw

                        }
```

c

```
# Random-effects NM Bayesian extension of Lumley's method

model{
        for (i in 1:N){
           prec[i]<- 1/var[i]
           diff[i]~dnorm(delta[i],prec[i])
           delta[i] ~ dnorm(md[i],taud[i])            # trial-specific means diff distributions
           taud[i] <- tau                             # precisions of diff distributions
           md[i] <- d[t[i]] - d[b[i]] +  w[comp[i]]}  # means of diff distributions

        # vague priors for basic parameters

        d[1]<-0
        for (k in 2:NT) {d[k] ~ dnorm(0,.00001) }

        # vague prior for RE standard deviation

        sd~dunif(0,10)
        tau<-1/pow(sd,2)
        tau2<-1/tau

        # All pairwise differences

        for (c in 1:NT-1){for (k in (c+1):NT){pairwise[c,k]<-d[k]-d[c]}}
        for (c in 1:Ncomp){w[c]~dnorm(0,tauw)}

        # vague prior for w

        sdw~dunif(0,10)
        tauw<-1/pow(sdw,2)
        tau2w<-1/tauw

                    }
```

d

```
# Random-effects NM Consistency model (Lu and Ades 2006) - Heterogeneous variance

model{
        for (i in 1:N){
           prec[i]<- 1/var[i]
           diff[i]~dnorm(delta[i],prec[i])
           delta[i] ~ dnorm(md[i],taud[i])       # trial-specific means diff distributions
           taud[i] <- tau * 1/exp(ni[b[i],t[i]]) # precisions of diff distributions
           md[i] <- d[t[i]] - d[b[i]]            # means of diff distributions
           for (c in 1:NT-1){for (k in (c+1):NT){ni[c,k]~dnorm(0,0.2)}}

        # vague priors for basic parameters

        d[1]<-0
        for (k in 2:NT) {d[k] ~ dnorm(0,.00001) }

        # vague prior for RE standard deviation

        sd~dunif(0,10)
        tau<-1/pow(sd,2)
        tau2<-1/tau


        # All pairwise differences

        for (c in 1:NT-1){for (k in (c+1):NT){pairwise[c,k]<-d[k]-d[c]}}

                    }
```

Fig. A3. WinBUGS codes for the (a) NM Consistency model, (b) NM Inconsistency model, (c) NM Bayesian extension of Lumley's method, (d) NM Heterogeneous model.

Appendix C: additional WinBUGS code for the NM models

The 6 interventions can be ranked in efficacy and the probability that each is best cab be obtained as follows:

a

```
# Ranking and prob{treatment k is best}
        for (k in 1:NT) {
                rk[k]<- NT + 1 - rank(d[],k)
                best[k]<-equals(rk[k],1)
        }
```

b

```
# Residual Deviance for data i
for (i in 1:N){
      dev[i] <- (diff[i]-delta[i])* (diff[i]-delta[i])/var[i]
      }
sumdev <- sum(dev[])
```

Fig. A4. WinBUGS codes for the (a) Ranking of treatments and Best probability; (b) Mean residual deviance.

Bibliography

References to studies included in the analyses

Aichelmann-Reidy ME, Yukna RA, Evans GH, Nasr HF & Mayer ET. (2001) Clinical evaluation of acellular allograft dermis for the treatment of human gingival recession. Journal of Periodontology 72, 998-1005.

Amarante ES, Leknes KN, Skavland J & Lie T. (2000) Coronally positioned flap procedures with or without a bioabsorbable membrane in the treatment of human gingival recession. Journal of Periodontology 71, 989-998.

Borghetti A, Glise J-M, Monnet-Corti V & Dejou J. (1999) Comparative clinical study of a bioabsorbable membrane and subepithelial connective tissue graft in the treatment of human gingival recession. Journal of Periodontology 70, 123-130.

Castellanos AT, de la Rosa MR, de la Garza M & Caffesse RG. (2006) Enamel matrix derivative and coronal flaps to cover marginal tissue recessions. Journal of Periodontology 77, 7-14.

Cortellini P, Baldi C, Francetti L, Raperini G, Rotundo R, Nieri M, Franceschi D, Labriola, A, Tonetti M & Pini Prato G. Comparative study of coronally advanced flap and bilaminar technique in the treatment of gingival recessions. A multi-centre, randomized, double blind, clinical study. Unpublished Data.

Córtes AQ, Martins AG, Nociti FH Jr, Salllum AW, Casati MZ & Sallum EA. (2004) Coronally positioned flap with or without acellular dermal matrix graft in the treatment of class I gingival recessions: a randomized controlled clinical study. Journal of Periodontology 75, 1137-1144.

Córtes AQ, Sallum AW, Casati MZ, Nociti FH Jr & Sallum EA. (2006) A two-year prospective study of coronally positioned flap with or without acellular dermal matrix graft. Journal of Clinical Periodontology 33, 683-689.

da Silva RC, Joly JC, de Lima AF & Tatakis DN. (2004) Root coverage using the coronally positioned flap with or without a subepithelial connective tissue graft. Journal of Periodontology 75, 413-419.

Del Pizzo M, Zucchelli G, Modica F, Villa R & Debernardi C. (2005) Coronally advanced flap with or without enamel matrix derivative for root coverage: a 2-year study. Journal of Clinical Periodontology 32, 1181-1187.

Huang L-H, Neiva REF, Soehren SE, Giannobile WV & Wang H-L. (2005) The effect of platelet-rich plasma on the coronally advanced flap root coverage procedure: a pilot human trial. Journal of Periodontology 76, 1768-1777.

Jepsen K, Heinz B, Halben JH & Jepsen S. (1998) Treatment of gingival recession with titanium reinforced barrier membranes versus connective tissue grafts. Journal of Periodontology 69, 383-391.

Joly JC, Carvalho AM, da Silva RC, Ciotti DL & Cury PR. (2007) Root coverage in isolated gingival recessions using autograft versus allograft: a pilot study. Journal of Periodontology 78, 1017-1022.

Leknes KN, Amarante ES, Price DE, Boe OE, Skavland RJ & Lie T. (2005) Coronally positioned flap procedures with or without a biodegradable membrane in the treatment of human gingival recession. A 6- year follow-up study. Journal of Clinical Periodontology 32, 518-529.

Lins LH, de Lima AF & Sallum AW. (2003) Root coverage: comparison of coronally positioned flap with or without titanium-reinforced barrier membrane. Journal of Periodontology 74, 168-174.

McGuire MK & Nunn M. (2003) Evaluation of human recession defects treated with coronally advanced flaps and either enamel matrix derivative or connective tissue. Part 1: Comparison of clinical parameters. Journal of Periodontology 74, 1110-1125.

Modica F, Del Pizzo M, Roccuzzo M & Romagnoli R. (2000) Coronally advanced flap for the treatment of buccal gingival recessions with and without enamel matrix derivative. A split-mouth study. Journal of Periodontology 71, 1693-1698.

Paolantonio M, Dolci M, Esposito P, D'Archivio D, Lisanti L, Di Luccio A & Perinetti G. (2002) Subpedicle acellular dermal matrix graft and autogenous connective tissue graft in the treatment of gingival recessions: a comparative 1-year clinical study. Journal of Periodontology 73, 1299-1307.

Pilloni A, Paolantonio M & Camargo PM. (2006) Root coverage with coronally positioned flap used in combination with enamel matrix derivative: 18-month clinical evaluation. Journal of Periodontology 77, 2031-2039.

Romagna-Genon C. (2001) Comparative clinical study of guided tissue regeneration with a bioabsorbable bilayer collagen membrane and subepithelial connective tissue graft. Journal of Periodontology 72, 1258-1264.

Spahr A, Hägewald S, Tsoulfidou F, Rompola E & Heijl L. (2005) Coverage of Miller class I and II recession defects using enamel matrix proteins versus coronally advanced flap technique: a 2-year report. Journal of Periodontology 76, 1871-1880.

Tal H, Moses O, Zohar R, Meir H & Nemcovsky C. (2002) Root coverage of advanced gingival recession: A comparative study between acellular dermal matrix allograft and subepithelial connective tissue grafts. Journal of Periodontology 73, 1405-1411.

Tatakis DN & Trombelli L. (2000) Gingival recession treatment: guided tissue regeneration with bioabsorbable membrane versus connective tissue graft. Journal of Periodontology 71, 299-307.

Trombelli L, Scabbia A, Tatakis DN & Calura G. (1998) Subpedicle connective tissue graft versus guided tissue regeneration with bioabsorbable membrane in the treatment of human gingival recession defects. Journal of Periodontology 69, 1271-1277.

Wang H-L, Bunyaratavej P, Labadie M, Shyr Y & MacNeil RL. (2001) Comparison of 2 clinical techniques for treatment of gingival recession. Journal of Periodontology 72, 1301-1311.

Wilson TG Jr, McGuire MK & Nunn ME. (2005) Evaluation of the safety and efficacy of periodontal applications of a living tissue-engineered human fibroblast-derived dermal substitute. II. Comparison to the sub-epithelial connective tissue graft: a randomized controlled feasibility study. Journal of Periodontology 76, 881-889.

Additional references

Allen EP & Miller PD. (1989) Coronal positioning of existing gingiva: short term results in the treatment of shallow marginal tissue recession. Journal of Periodontology 60, 316-319.

American Academy of Periodontology (1996) Consensus report on Mucogingival Therapy. Proceedings of the World Workshop in Periodontics. Annals of Periodontology 1, 702-706.

Baayen RH, Davison DJ & Bates DM. (2008) Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language 59, 390-412.

Berkey CS, Anderson JJ, Hoaglin DC. (1996) Multiple-outcome meta-analysis of clinical trials. Stat Med 15:537-57.

Berkey CS, Hoaglin DC, Antczak-Bouckoms A, et al. (1998) Meta-analysis of multiple outcomes by regression with random effects. Stat Med 17:2537-50.

Cheng YF, Chen JW, Lin SJ & Lu HK. (2007) Is coronally positioned flap procedure adjunct with enamel matrix derivative or root conditioning a relevant predictor for achieving root coverage? A systematic review. Journal of Periodontal Research 42, 474-485.

Chung H, Lumley T. (2008) Graphical exploration of network meta-analysis data: the use of multidimensional scaling. Clinical Trials 5: 301-307.

Clauser C, Nieri M, Franceschi D, Pagliaro U & Pini-Prato G. (2003) Evidence-based mucogingival therapy. Part 2: ordinary and individual patient data meta-analyses of surgical treatment of recession using complete root coverage as the outcome variable. Journal of Periodontology 74, 741-756.

Domenici F, Parmigiani G, Wolpert RL, Hasselblad V. (1999) Meta-analysis of migraine headache treatments: combining information from heterogenous designs. Journal of the American Statistical Association 94:16-28.

Elliott WJ, Meyer PM. (2007) Incident diabetes in clinical trials of antihyperintensive drugs: network meta-analysis. Lancet 369:201-7.

Deeks JJ, Higgins JPT, Altman DG. (2008) Analyzing data and undertaking meta-analyses. In: Higgins JPT, Green S, eds., Cochrane Handbook of Systematic Reviews of Interventions. Chichester, England: John Wiley & Sons.

Dias S, Welton NJ, Caldwell DM, et al. (2010) Checking consistency in mixed treatment comparison meta-analysis. Stat Med 29:932-44.

Gelman A, and Rubin RD (1992). Inferences From Iterative Simulation Using Multiple Sequences. Statistical Science 7, 457-472.

Gelman A. (2006) Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 3:515-533.

Harris RJ. (1998) Root coverage with a connective tissue with partial thickness double pedicle graft and an acellular dermal matrix graft: a clinical and histological evaluation of a case report. Journal of Periodontology 69, 1305-1311.

Hasselblad V. (1998) Meta-analysis of multi-treatment studies. Medical Decision Making 18:37-43.

Higgins JPT, Green S (eds). (2005) Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: Wiley.

Kullback S and Leibler RA, Ann. Math. Stat. 22, 79 (1951); Kullback S, Information Theory and Statistics (Wiley, New York, 1959).

Lambert PC, Sutton AJ, Burton PR, et al. (2005) How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Stat Med 24:2401-28.

Lu G, Ades AE, Sutton AJ, et al. (2007) Meta-analysis of mixed treatment comparisons at multiple follow-up times. Stat Med 26:3681-99.

Marx RE, Carlson ER, Eichstaedt RM, Schimmele SR, Strauss JE & Georgeff KR (1998) Platelet-rich plasma: growth factor enhancement for bone grafts. Oral Surgery, oral medicine, oral pathology, oral radiology, and endodontics 85, 638-646.

McCulloch R. Local model influence. (1989) J Am Statistical Assoc 84:473-8.

Miller PD. (1985) A classification of marginal tissue recession. International Journal of Periodontics and Restorative Dentistry 5, 8-13.

Nam I-S, Mengerson K, Garthwaite P. (2003) Multivariate meta-analysis. Stat Med 22:2309-33.

Oates, TW, Robinson M & Gunsolley JC. (2003) Surgical therapies for the treatment of gingival recession. A systematic review. Annals of Periodontology 8, 303-320.

Pini Prato G, Tinti C, Vincenzi G, Magnani C, Cortellini P & Clauser C. (1992) Guided tissue regeneration versus mucogingival surgery in the treatment of human buccal gingival recession. Journal of Periodontology 63, 919-928.

Pini Prato G, Pagliaro U, Baldi C, Nieri M, Saletta D, Cairo F & Cortellini P. (2000) Coronally advanced flap procedure for root coverage. Flap with tension versus flap without tension: a randomized controlled clinical study. Journal of Periodontology 71, 188 201.

Rasperini G, Silvestri M, Schenk RK & Nevins ML. (2000) Clinical and histologic evaluation of human gingival recession treated with a subepithelial connective tissue graft and enamel matrix derivative (Emdogain): a case report. International Journal of Periodontics and Restorative Dentistry 20, 269-275.

Riley RD, Abrams KR, Lambert PC, et al. (2007) An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. Stat Med 26:78-97.

Roccuzzo M, Bunino M, Needleman I & Sanz M. (2002) Periodontal plastic surgery for treatment of localized gingival recessions: a systematic review. Journal of Clinical Periodontology 29, 178-194.

Spiegelhalter DJ, Thomas A, Best NG, and Gilks, WR (1996a). BUGS Examples Volume 1, Version 0.5. Cambridge: MRC Biostatistics Unit. - (1996b). BUGS Examples Volume 2, Version 0.5. Cambridge: MRC Biostatistics Unit.

Spiegelhalter DJ, Best NG, Carlin BP, et al. (2002a) Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society (B) 64:583-616.

Sutton AJ, Abrams KR. (2001) Bayesian methods in meta-analysis and evidence synthesis. Stat Methods Med Res 10:277-303.

Sutton A, Ades AE, Cooper N & Abrams K. (2008) Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics 26, 753-767.

Turner RM, Omar RZ, and Thompson SG (2001). Bayesian methods of analysis for cluster randomized trials with binary outcome data. Statistics in Medicine 20: 453-472. 543

van der Valk R, Webers CAB, Lumley T, Hendrikse F, Prins MH, Schouten JSAG. (2009) A network meta-analysis combined direct and indirect comparisons between glaucoma drugs to rank effectiveness in lowering intraocular pressure. Journal of Clinical Epidemiology 62(12):1279-83.

van Houwelingen HC, Arends LR & Stijnen T. (2002) Advanced methods in meta-analysis: multivariate approach and meta-regression. Statistics in Medicine 21, 589-624.

Wennström JL & Zucchelli G. (1996) Increased gingival dimensions. A significant factor for successful outcome of root coverage procedures? A 2-year prospective clinical study. Journal of Clinical Periodontology 23, 770-777.

Whitehead A. (2002) Meta-analysis of controlled clinical trials. Chichester, UK.