Diogo Manuel de Jesus Mendes

BENEFIT-RISK ASSESSMENT OF MEDICINES: A CONTRIBUTION FOR EVIDENCE-BASED DECISIONS USING THE NUMBER NEEDED TO TREAT TO BENEFIT OR TO HARM

Tese de doutoramento em Ciências Farmacêuticas, especialidade de Farmácia Clínica, orientada pelo Professor Doutor Francisco Jorge Batel Marques e apresentada à Faculdade de Farmácia da Universidade de Coimbra

Junho 2017

· U C ·

Universidade de Coimbra

Diogo Manuel de Jesus Mendes

Benefit-risk assessment of medicines: a contribution for evidence-based decisions using the number needed to treat to benefit or to harm

Tese de doutoramento em Ciências Farmacêuticas, especialidade de Farmácia Clínica, orientada pelo Professor Doutor Francisco Jorge Batel Marques e apresentada à Faculdade de Farmácia da Universidade de Coimbra

Junho 2017

UNIVERSIDADE DE COIMBRA

All the research work presented in this thesis was performed in strict collaboration of the Laboratory of Pharmacology, Faculty of Pharmacy, University of Coimbra and the Centre for Health Technology Assessment and Drug Research, Association for Innovation and Biomedical Research on Light and Image, under the supervision of Professor Francisco Jorge Batel Marques.

Todos os trabalhos experimentais apresentados nesta tese foram realizados em estreita colaboração entre o Laboratório de Farmacologia da Faculdade de Farmácia da Universidade de Coimbra e o Centro de Avaliação de Tecnologias em Saúde e Investigação do Medicamento da Associação para Investigação Biomédica e Inovação em Luz e Imagem, sob a supervisão do Professor Doutor Francisco Jorge Batel Marques.

Aos meus pais.

AGRADECIMENTOS

Ao meu orientador, Professor Doutor Francisco Batel Marques, agradeço pela orientação científica, disponibilidade, profissionalismo e humanismo. Agradeço também o convite que me endereçou para realizar este trabalho. O facto de nunca ter deixado de acreditar no nosso trabalho levou-o a bom porto. Agradeço ainda ao Professor Doutor Francisco Batel Marques por me integrar como colaborador da Unidade de Farmacovigilância de Coimbra (UFC) e do Centro de Avaliação de Tecnologias em Saúde e Investigação do Medicamento (CHAD), da AIBILI – Associação para a Investigação Biomédica e Inovação em Luz e Imagem. Esta interação acrescentou um enorme valor a este trabalho. E também pelo seu companheirismo e confiança pessoal.

Ao Professor Doutor Carlos Miguel Costa Alves, da Faculdade de Farmácia da Universidade de Coimbra e do CHAD, pelo seu contributo para a condução dos trabalhos, análise crítica, amizade e companheirismo.

À Dra. Ana Penedones, do CHAD, que nestes últimos meses assegurou a realização de uma parte considerável das tarefas da UFC.

À AIBILI – Associação para a Investigação Biomédica e Inovação em Luz e Imagem por me ter apoiado inteiramente nos meus trabalhos. É uma honra colaborar com uma instituição com o reconhecimento científico nacional e internacional como o da AIBILI.

À Faculdade de Farmácia da Universidade de Coimbra por me ter proporcionado condições para conduzir este trabalho.

À minha família, em particular aos meus pais, irmão e à Sílvia pelo apoio incondicional.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVIATIONS

LIST OF ABREVIATIONS

^{13}C-UBT	^{13}C Urea Breath Test
ACS	Acute Coronary Syndrome
ADVANCE	Accelerated Development Of Vaccine Benefit-Risk Collaboration In Europe
AE	Adverse Event
AELD	Adverse Event Leading To Discontinuation Of Study Drug
AE-NNT	Adverse Event Adjusted Number Needed To Treat
AIDS	Acquired Immune Deficiency Syndrome
AFFIRM	Natalizumab Safety And Efficacy In Relapsing-Remitting Multiple Sclerosis
AHEAD	Action For Health In Diabetes
AHF	Acute Heart Failure
ALT	Alanine Transaminase
AMI	Acute Myocardial Infarction
APPROVe	Adenomatous Polyp Prevention On Vioxx
AR	Absolute Risk
ARI	Absolute Risk Increase
ARR	Annualized Relapse Rate
ASF	Ashby And Smith Framework
BCC	Blood Cell Counts
BEYOND	Betaferon/Betaseron Efficacy Yielding Outcomes Of A New Dose In Multiple Sclerosis Patients
BID	Twice A Day
BLRA	Benefit-Less–Risk Analysis
BMJ	British Medical Journal
BRAFO	Benefit–Risk Analysis For Foods
BRAT	Benefit-Risk Action Team
BR	Benefit-Risk
BRR	Benefit–Risk Ratio
CA	Conjoint Analysis
CAD	Coronary Artery Disease
CARE-MS	Comparison Of Alemtuzumab And Rebif Efficacy In Multiple Sclerosis
CASS	Taskforce Of Representatives From Health Canada, Australia's

	Therapeutic Goods Administration, Swissmedic, And The Singapore Health Science Authority
CCA	Carlos Costa Alves
CDPS3M	Confirmed Disability Progression Sustained For 3 Months
CDPS6M	Confirmed Disability Progression Sustained For 6 Months
CDS	Cross-Design Synthesis
CENTRAL	Cochrane Central Register Of Controlled Trials
CER	Control Event Rate
CHD	Coronary Heart Disease
CHF	Congestive Heart Failure
CI	Confidence Interval
CIOMS	Council For International Organizations Of Medical Sciences
CIRS	Centre For Innovation In Regulatory Science
CMR-CASS	Centre for Medicines Research Health Canada, Australia's Therapeutic Goods Administration, SwissMedic and Singapore Health Science Authority
COBRA	Consortium On Benefit-Risk Assessment
COPD	Chronic Obstructive Pulmonary Disease
CONFIRM	Comparator And An Oral Fumarate In Relapsing–Remitting Multiple Sclerosis
CoV	Contingent Valuation
CPM	Confidence Profile Method
CRV	Coronary Revascularization
CUI/DI	Clinical Utility Index/Desirability Index
CV	Cardiovascular
CVD	Cardiovascular Disease
DAG	Directed Acyclic Graphs
DALY	Disability-Adjusted Life Years
DAPT	Dual Antiplatelet Therapy
DCE	Discrete Choice Experiment
DEFINE	Determination Of The Efficacy And Safety Of Oral Fumarate In Relapsing-Remitting Multiple Sclerosis
DILI	Drug-Induced Liver Injury
DMF	Dimethyl Fumarate

DMM	Diogo Manuel Mendes
DMT	Disease-Modifying Therapy
DVT	Deep Venous Thrombosis
EBM	Evidence-Based Medicine
EDSS	Expanded Disability Status Scale
EIN	Exposure Impact Number
EMA	European Medicines Agency
EOD	Every Other Day
EPAR	European Public Assessment Report
ERN	Nicotinic Acid
EU	European Union
EXTEND	Extended Prophylactic Treatment With Melagatran/Ximelagatran Versus Enoxaparin for the Prevention of Venous Thromboembolic Events in Patients Undergoing Elective Hip Replacement or Hip Fracture Surgery
FBM	Francisco Batel Marques
FDA	Food And Drug Administration
FDA BRF	Food And Drug Administration Benefit-Risk Framework
FREEDOMS	Fingolimod Research Evaluating Effects Of Daily Oral Therapy In Multiple Sclerosis
GA	Glatiramer Acetate
GBR	Global Benefit–Risk
GGT	Gamma-Glutamyl Transferase
GOT	Glutamic Oxaloacetic Transaminase
GPT	Glutamate-Pyruvate Transaminase
HALE	Health-Adjusted Life Years
HbA1c	Glycated Haemoglobin
HDL	High-Density Lipoprotein
HPS2-THRIVE	Heart Protection Study 2 – Treatment Of HDL To Reduce The Incidence Of Vascular Events
HR	Hazard Ratio
HVD	Heart Valve Disease
IAR	Infusion-Associated Reaction
ICU	Intensive Care Unit
IFN	Interferon

IM	Intramuscular
IMI	Innovative Medicine Initiative
INHB	Incremental Net Health Benefit
IPIR	Immediate Post-Injection Reaction
ISR	Injection-Site Reaction
ITC	Indirect Treatment Comparison
ITP	Immune Thrombocytopenic Purpura
IV	Intravenous
JCV	John Cunningham Virus
LDL	Low-Density Lipoprotein
LHH	Likelihood To Be Helped Or Harmed
LRPT	Laropiprant
LRTI	Low Respiratory Tract Infection
MA	Marketing Authorization
MACE	Major Adverse Cardiovascular Event
MAH	Marketing Authorization Holder
MAR	Maximum Acceptable Risk
MCDA	Multi-Criteria Decision Analysis
MCE	Minimum Clinical Efficacy
MDP	Markov Decision Process
MedDRA	Medical Dictionary For Regulatory Activities
MEDLINE	Medical Literature Analysis And Retrieval System Online
MET	Metformin
MI	Myocardial Infarction
Mo	Month
MP	Myopathy
MRI	Magnetic Resonance Imaging
MRP	Mutual Recognition Procedure
MSCRG	Multiple Sclerosis Collaborative Research Group
MTC	Mixed Treatment Comparison
NA	Not Available
NCB	Net Clinical Benefit
NEAR	Net Efficacy Adjusted For Risk
NEPP	Number Of Events Prevented In The Population

NMI	Nonfatal Myocardial Infarction
NNE	Number Needed To Be Exposed
NNEB	Number Needed To Be Exposed For One Person To Benefit
NNEH	Number Needed To Be Exposed For One Person To Be Harmed
NNH	Number Needed To Harm
NNT	Number Needed To Treat
NNTB	Number Needed To Treat To Benefit
NNTH	Number Needed To Treat To Harm
NOD	New-Onset Diabetes
NR	Not Reported
NSAID	Non-Steroidal Anti-Inflammatory Drug
OCD	Obsessive-Compulsive Disorder
OHA	Oral Hypoglycaemic Agent
OMERACT 3x3	Outcome Measures In Rheumatology 3 × 3
OR	Odds Ratio
ORR	Objective Response Rate
OS	Overall Survival
PAES	Post-Authorization Efficacy Studies
PASS	Post-Authorization Safety Studies
PCI	Percutaneous Coronary Intervention
PCP	Pneumocystis Pneumonia
PD	Psychiatric Disorders
PE	Pulmonary Embolism
PEER	Patient's Expected Event Rate
PhRMA	Pharmaceutical Research And Manufacturers Of America
PIN-ER-*t*	Population Impact Number Of Eliminating A Risk Factor Over Time t
PIO	Pioglitazone
PFS	Progression-Free Survival
PML	Progressive Multifocal Leukoencephalopathy
PN	Peripheral Neuropathy
PP-FCDPS3M	Proportion Of Patients Remaining Free Of Confirmed Disability Progression Sustained For 3 Months
PPI	Proton-Pump Inhibitor
PPR	Proportion Of Patients With Relapse

PPR-F	Proportion Of Patients Remaining Relapse-Free
PRISMA	Preferred Reporting Items For Systematic Reviews And Meta-Analyses
PRISMS	Prevention Of Relapses And Disability By Interferon Beta-1a Subcutaneously In Multiple Sclerosis
PRO	Patient Reported Outcome
PROACTIVE	Prospective Pioglitazone Clinical Trial In Macrovascular Events
PrROACT-URL	Problem, Objectives, Alternatives, Consequences, Trade-Offs, Uncertainty, Risk And Linked Decisions
PSM	Probabilistic Simulation Method
PY	Person-Year
QALY	Quality-Adjusted Life Year
QoL	Quality Of Life
Q-TWiST	Quality-Adjusted Time Without Symptoms And Toxicity
RA	Regulatory Authority
RCT	Randomized Controlled Trial
RD	Risk Difference
RECIST	Response Evaluation Criteria In Solid Tumours
RECORD	Rosiglitazone Evaluated For Cardiac Outcomes And Regulation Of Glycaemia In Diabetes
REGULATE	Randomised, Double-Blind Study With Comparison Of Benfluorex Versus Pioglitazone In Combination With Sulfonylurea Administered Orally For The Treatment Of Type 2 Diabetes
RET	Rearranged During Transfection
RMP	Risk Management Plan
RMS	Risk Management System
RR	Relative Risk
RRI	Relative Risk Increase
RRR	Relative Risk Reduction
RRMS	Relapsing–Remitting Multiple Sclerosis
RSG	Rosiglitazone
RV-MCE	Relative Value-Adjusted Minimum Clinical Efficacy
RV-NNH	Relative Value-Adjusted Number Needed To (Treat To) Harm
SABRE	Southeast Asia Benefit-Risk Evaluation
SAE	Serious Adverse Event

SB	Serious Bleeding
SBRAM	Sarac's Benefit–Risk Assessment
SC	Subcutaneous
SCOUT	Sibutramine Cardiovascular Outcomes Trial
SI	Serious Infection
SMAA	Stochastic Multi-Criteria Acceptability Analysis
SmPC	Summary Of Product Characteristics
SOC	System Organ Class
SPM	Stated Preference Method
SPORTIF	Stroke Prevention Using Oral Thrombin Inhibitor In Atrial Fibrillation
SSRI	Selective Serotonin Reuptake Inhibitor
STRADIVARIUS	Strategy To Reduce Atherosclerosis Development Involving Administration Of Rimonabant – The Intravascular Ultrasound Study
SU	Sulfonylurea
TEMSO	Teriflunomide Multiple Sclerosis Oral
TIA	Transient Ischemic Attack
TMF	Tumour Necrosis Factor
TIW	Thrice A Week
TURBO	Transparent Uniform Risk–Benefit Overview
TZD	Thiazolidinedione
UER	Unexposed Event Rate
ULN	Upper Limit Of Normal
UMBRA	Unified Methodology For Benefit-Risk Assessment
USA	United States Of America
UT-NNT	Utility-Adjusted And Time-Adjusted Number Needed To Treat
VIGOR	Vioxx Gastrointestinal Outcomes Research Trial
WHO	World Health Organization
Wk	Week
wNCB	Weighted Net Clinical Benefit
WP	Work Programme
WSMI BRAND	World Self Medication Industry Benefit-Risk Assessment For Non-Prescription Drugs
Yr	Year

PUBLICATIONS

PUBLICATIONS

Full text publications

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2015. NUMBER NEEDED TO HARM IN THE POST-MARKETING SAFETY EVALUATION: RESULTS FOR ROSIGLITAZONE AND PIOGLITAZONE. PHARMACOEPIDEMIOL DRUG SAF, 24(12), PP. 1259-1270.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2016A. TESTING THE USEFULNESS OF THE NUMBER NEEDED TO TREAT TO BE HARMED (NNTH) IN BENEFIT-RISK EVALUATIONS: CASE STUDY WITH MEDICINES WITHDRAWN FROM THE EUROPEAN MARKET DUE TO SAFETY REASONS. EXPERT OPIN DRUG SAF, 15(10), PP. 1301-1312.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2016B. BENEFIT-RISK OF THERAPIES FOR RELAPSING-REMITTING MULTIPLE SCLEROSIS: TESTING THE NUMBER NEEDED TO TREAT TO BENEFIT (NNTB), NUMBER NEEDED TO TREAT TO HARM (NNTH) AND THE LIKELIHOOD TO BE HELPED OR HARMED (LHH): A SYSTEMATIC REVIEW AND META-ANALYSIS. CNS DRUGS, 30(10), PP. 909-929.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2017. NUMBER NEEDED TO TREAT (NNT) IN CLINICAL LITERATURE: AN APPRAISAL. BMC MEDICINE, DOI: 10.1186/S12916-017-0875-8.

ABSTRACT/RESUMO

ABSTRACT

The assessment of benefit-risk ratios is performed during the entire life-cycle of medicines. Although pre-marketing randomized controlled clinical trials provide important evidence about benefits and harms of medicines in well-defined populations, they have few intrinsic limitations. Those studies are usually underpowered to detect rare and/or long term latency adverse events. Then, serious safety problems may be identified only during post-marketing, namely by means of observational studies or spontaneous reporting schemes.

Regulatory authorities closely monitor post-marketing safety signals, aiming to assure that only medicines with favourable benefit-risk ratios are available for use. When regulatory authorities conclude that the benefits no longer outweigh the risks of a medicine, they initiate actions in order to protect public health. There are several examples of medicines that were recently suspended or withdrawn from market because of safety problems. However, different regulatory authorities may reach divergent conclusions about the benefit-risk ratio of a medicine despite analysing the same evidence. One of the reasons may be the fact that the assessments still rely heavily on expert opinions and subjective qualitative weighing of the available evidence. Therefore, several projects have been initiated aiming at testing and developing methodologies that could potentially bring clarity to the decision-making process and help regulatory authorities to make more objective, consistent and evidence-based decisions. The introduction of structured frameworks that could encompass quantitative or semi-quantitative methodologies was advocated. The number needed to treat (NNT) is one of the methodologies recommended for testing in benefit-risk assessments.

This project was carried out to evaluate the potential usefulness of the NNT as quantitative metric for post-marketing benefit-risk assessment of medicines, using several case studies and addressing both regulatory and clinical perspectives. There is limited evidence about the usefulness of this metric to support regulatory decisions on benefit-risk assessment.

The NNT can be effectively used to quantity benefits and risks of medicines, as well as to provide additional and useful information about the magnitude of treatment effects. From a regulatory perspective, the use of the NNT may be considered only within defined structured frameworks for benefit-risk assessment, because there are several issues weighing in the assessments that are not addressed by quantitative metrics. The application of the NNT can be problematic for weighing multiple benefits and risks with different clinical relevance. Nonetheless, whenever calculable, the NNT may be used in the benefit-risk assessment of medicines, as this metric can help to strengthen regulatory decisions. In

addition, the NNT is useful for supporting informed clinical decision-making, as long as it is properly calculated. In conclusion, although the NNT does not replace other evaluations in the benefit-risk assessment of marketed medicines, it provides useful information, as well as added value in well-defined assessments.

RESUMO

A avaliação da relação benefício-risco é realizada durante todo o ciclo de vida do medicamento. Apesar dos ensaios clínicos aleatorizados e controlados fornecerem evidência acerca dos benefícios e riscos dos medicamentos em populações bem definidas, estes estudos apresentam algumas limitações intrínsecas, incluindo o facto de não apresentarem, frequentemente, o poder estatístico necessário para a deteção de eventos adversos raros e/ou de longo tempo de latência. Como tal, podem ocorrer problemas graves de segurança que são identificados apenas durante a fase de pós-comercialização, designadamente através de estudos observacionais ou sistemas de notificação espontânea.

As autoridades reguladoras monitorizam de forma cuidadosa os sinais de segurança gerados durante a fase de pós-comercialização, tendo por objetivo assegurar que apenas os medicamentos com relação benefício-risco positiva continuem disponíveis para utilização. Caso concluam que os benefícios de um medicamento deixaram de superar os seus riscos, as autoridades reguladoras desencadeiam ações regulamentares tendo em vista a proteção da saúde pública. Existem vários exemplos de medicamentos que foram recentemente suspensos ou retirados do mercado devido a problemas de segurança. No entanto, diferentes autoridades reguladoras podem chegar a conclusões diferentes acerca da relação benefício-risco de um medicamento, apesar de analisarem a mesma evidência. Assim, têm sido iniciados vários projetos destinados a testar e desenvolver metodologias que possam trazer clareza ao processo de decisão e auxiliar as autoridades reguladoras a fazerem decisões mais objetivas, consistentes e baseadas na evidência. Tem sido proposta a introdução de processos estruturados que possam incluir metodologias quantitativas. Uma das metodologias recomendadas para investigação no contexto da avaliação da relação benefício-risco de medicamentos é o número necessário tratar (NNT).

Este projeto foi desenvolvido para avaliar a utilidade do NNT como métrica quantitativa na avaliação pós-comercialização da relação benefício-risco de medicamentos, recorrendo a vários casos de estudo e com foco nas perspetivas regulamentar e clínica. Existe evidência limitada acerca da utilidade desta métrica como instrumento de suporte à tomada de decisões acerca da relação benefício-risco de medicamentos.

O NNT pode ser utilizado para quantificar benefícios e riscos, bem como para fornecer informações complementares úteis acerca da magnitude dos efeitos de um tratamento. Da perspetiva regulamentar, a utilização do NNT pode ser considerada apenas como uma parte integrante de processos estruturados e bem definidos para a avaliação de

relações benefício-risco, uma vez que existem vários problemas com peso na avaliação que não são facilmente apreciados por estas métricas. A aplicação do NNT pode ser problemática quando é necessário avaliar múltiplos benefícios e riscos com relevâncias clínicas diferentes. No entanto, sempre que seja possível calcular, o NNT pode ser utilizado na avaliação das relações risco-benefício de medicamentos, uma vez que esta métrica pode ajudar a reforçar as decisões regulamentares. Além disso, o NNT é útil para apoiar a decisão clínica informada, desde que seja devidamente calculado. Em conclusão, embora o NNT não substitua outras avaliações na avaliação das relações benefício-risco de medicamentos comercializados, esta métrica fornece informações úteis, bem como valor acrescentado em avaliações bem definidas.

CHAPTER I – GENERAL INTRODUCTION

I. GENERAL INTRODUCTION

I.1. BENEFIT-RISK OF MEDICINES

I.1.1. RISK AVERSION VERSUS PUBLIC HEALTH

Regulatory authorities are responsible for ensuring that all medicines marketed under their supervision are safe, effective and of high quality. In the European Union (EU), the European Medicines Agency (EMA) aims to facilitate the development and access to medicines, to evaluate applications for marketing authorization purposes, to monitor the safety of medicines across their life cycle, and to provide information on medicines to healthcare professionals and patients (EMA 2016a).

In the context of marketing authorization, only medicines for which the benefits outweigh risks are licensed. Benefit-risk assessment is a key component in regulatory decision making about licensing medicines for use in a given patient population (EMA 2015a). However, pre-marketing studies are usually insufficient to fully characterize the benefit-risk profile of new medicines, with resulting uncertainties around the size of the effect and the probability of harm (Eichler et al. 2013). Thus, regulatory authorities face the challenge of balancing early patient access to new medicines with the need for more and better data on the benefits and risks of those medicines (Eichler et al. 2008).

This mission has become even more challenging in face of several safety alerts and medicinal product recalls during post-marketing in recent years. In the EU, 27 medicines were suspended or withdrawn from the market due to safety reasons between 2001 and 2015 (Mendes, Alves & Batel-Marques 2016a). In such environment, two points of view may arise. On one side, regulatory authorities have been criticized for being excessively risk-tolerant and accepting too many uncertainties when allowing medicines on the market (Eichler et al. 2008), (Garattini & Bertele 2007), (Carpenter, Zucker & Avorn 2008). On the other side, there are criticisms pointing out that regulatory authorities are overly risk-averse and request too much data before approving a medicine, highlighting the humanitarian cost of postponing or inhibiting access to potentially life-saving medicines (Eichler et al. 2008), (Eichler et al. 2013).

While the negative consequences of risk tolerance in approving medicines that cause significant harms are directly perceived, the adverse effects on public health due to the lack

of new medicines because of risk-aversion are often overlooked (Eichler et al. 2013). However, following the concept "first, do no harm" (the precautionary principle) in a rigid way, may not be the best option with regards to the interest of public health (Eichler et al. 2013). This principle fitted well years ago when there were few effective, but many potentially fatal, treatments (Lenert, Markowitz & Blaschke 1993). Today, there are treatments that are both effective and potentially harmful, such as those for life-threatening diseases. Thus, renouncing benefits because of a particular risk may lead to other risks, as those related to the progression of serious illness, with negative consequences on public health.

Figure I. 1 illustrates the relationship between the degrees of risk tolerance (or risk aversion) by regulatory authorities and the gains for public health owing to the development of medicines.

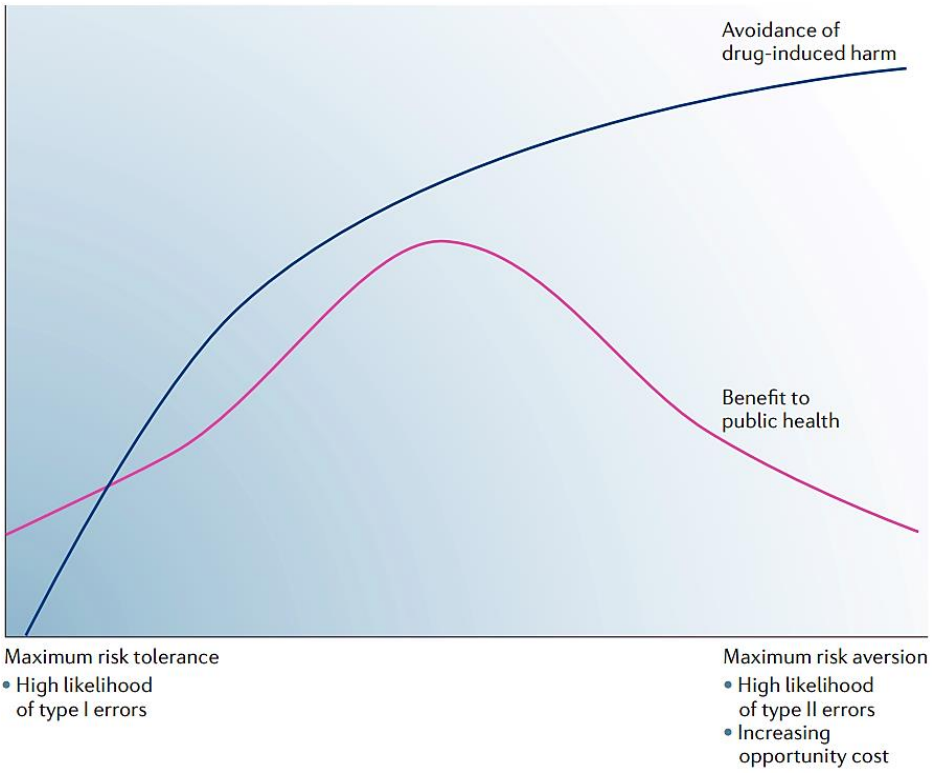


Figure I. 1 – Risk tolerance and risk aversion in drug regulation versus benefits to public health.

Risk tolerance and risk aversion by regulatory authorities (*x* axis) versus expected outcomes (*y* axis) in terms of avoidance of drug-induced harm (blue line) or net public health gains (purple line).
Source: (Eichler et al. 2013).

As presented in Figure I. 1, neither too much risk-tolerance, nor extreme risk-aversion are in the interest of public health. The first approach leads to the approval of unsafe or ineffective medicines (type I regulatory errors), while the latter precludes patients from receiving potentially important treatments (type II regulatory errors). Thus, regulatory

authorities should try to place themselves in the middle ground, attempting to maximize gains for public health through balancing benefits of treatment against the risks of treatment and the risks of untreated disease (Eichler et al. 2013).

The risk-risk trade-off concept, in which the risk from untreated disease is balanced against the risk of the treatment itself (Graham & Wiener 1995), is also a valid approach for making regulatory decisions. This concept may be useful to understand the influence of risk aversion on the decision-making process at various levels, depending on value judgments from different stakeholders. As an example, physicians tend to accept better death as a consequence of a disease rather than as an adverse reaction to a medicine. A survey-based study found that four or five lives have to be saved by treatment of the disease for each additional death caused by the treatment itself (Lenert, Markowitz & Blaschke 1993). Yet, patients seem to be more tolerant to iatrogenic risks than physicians (Johnson et al. 2010), (Byun et al. 2016). Other study used a discrete choice approach to compare benefit and risk preferences of regulators with those of physicians and diabetic patients. The authors concluded that the three stakeholders exhibited similar preferences concerning major effects (e.g. cardiovascular risk reduction as a favourable effect; persistent gastrointestinal problems as unfavourable effect), but attached different values to minor or short-term effects, with patients giving a higher importance to symptomatic adverse reactions, such as hypoglycaemia, as compared to regulators (Mol et al. 2015).

Based on the examples provided above, it seems unlikely that the various stakeholders agree on the outcome of the assessment of benefit-risk ratios of medicines in all cases. The tolerance of risk threshold is expected to depend on the context, as well as on the perspective of the assessor.

The authors of a study about the attitude of European regulators concluded that they may be perceived as risk averse, but differences between regulators and other stakeholders were not tested (EMA 2012). Such attitude is somehow understandable taking into account that regulatory authorities seem to be more often criticized for being excessively tolerant, allowing potentially unsafe medicines to reach the market, than for being risk averse (Jüni et al. 2004), (Nissen & Wolski 2010). Therefore, regulators may tend to place themselves on the risk aversion side, i.e. rejecting the approval of medicines, when there are uncertainties (Eichler et al. 2013).

The case of natalizumab provides an example of less risk tolerance from regulators as compared with patients. Natalizumab was introduced in the USA in 2004 for treating patients with relapsing-remitting multiple sclerosis (RRMS), after receiving approval from the

Food and Drug Administration (FDA). Within few months, three cases of progressive multifocal leukoencephalopathy (PML) were reported and the manufacturer decided, with support from the FDA, to voluntarily withdraw natalizumab from the market (FDA 2005). Natalizumab was reintroduced in the market in 2006 upon the recommendation of a FDA advisory board committee and at the request of patients (Kang 2006). According to the patients' opinion, they would prefer to receive a treatment that was more effective than the others available at the time, despite a chance of one in a thousand of suffering a fatal adverse event (Calfee 2006).

As with regulatory decisions in other fields of society, the decision of licensing a medicine is made under circumstances of uncertainty. Nevertheless, the level of acceptable uncertainty about benefits and risks is subject of debate (Moore & Furberg 2012). Yet, it is important to recall that efficacy and safety data obtained from pre-clinical studies are not always verified in the real-world background (Eichler et al. 2011). In addition, requesting too much data before approvals may have negative effects on public health (Figure I. 1), and lead to increasing amounts of investment (cost of opportunity) that in some situations produce small gains in knowledge (Eichler, et al., 2013). Further, benefits and risks change across the lifecycle of medicines.

I.1.2. MEANING OF BENEFIT AND RISK

It is of utmost importance to understand the meaning of benefit and risk for assessing benefit-risk of medicines. In the context of the first work package of the "Benefit-risk methodology project", members from the EMA project team visited five European regulatory authorities (Sweden, France, The Netherlands, UK and Spain) with the aim to gather knowledge about benefit-risk assessments practices (EMA 2011a).

One of the goals was to understand the meaning of benefit and risk for those regulators. The major conclusion was that the meanings are very fluid (EMA 2011a). The definitions provided by regulators varied across interviewees. Of note, more varied definitions were provided for risk than for benefit.

Most interviewees agreed that benefits are "clinically meaningful improvements to a patient, an improvement in health state or quality of life". Yet, other definitions were given: "improvement over a placebo, or at least non-inferior to comparators; a statistically significant effect; a change in the disease management of a patient; a better way of delivering a drug; or even a safety improvement" (EMA 2011a).

Interviewees found it more difficult to define risk as compared to benefit. They considered that benefits are objective and risks are not, being therefore more challenging to define. A consensual definition was not found. Several meanings were attributed to risk: "absence of benefit; dangers/hazards for the patient, adverse events, direct or indirect harm to the patient, frequency and severity of a side effect; harm to non-patients and to the general public; unacceptable damage to the patient; what is lost compared to current therapy; the negative aspects of a drug; the inverse of safety; pharmacokinetic interactions; insufficient duration; probability of an adverse event or harm; negative impact on quality of life; failure to meet endpoints; intolerability; uncertainty surrounding the risks; mortality; a concept of gambling which includes perception" (EMA 2011a).

According to the EMA project team, "from a decision-theoretic perspective, any drug decision could be decomposed into two broad components: Firstly, the favourable ("good things") or unfavourable ("bad things") effects for the patient; secondly, the level of uncertainty surrounding each of them" (EMA 2011a). These two aspects are illustrated in Figure I. 2, where the first column represents the values and the second column represents the associated uncertainties.

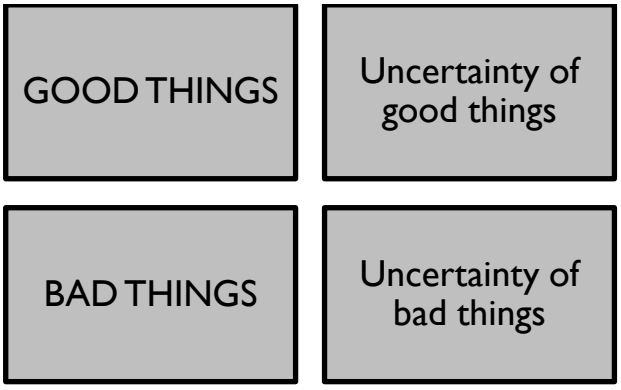


Figure I. 2 – Representation of effects and uncertainties around effects in the context of benefit-risk assessment of medicines.

The interviewees were further asked to allocate 100 points to each quadrant to reflect the time and effort spent at analysing each component. The distribution of points in each of the four quadrants is shown Figure I. 3.

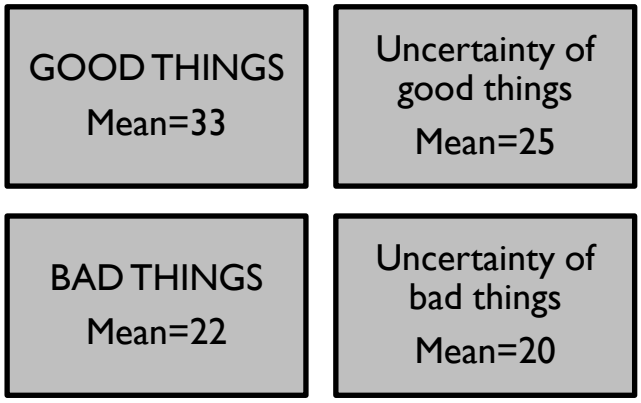


Figure I. 3 – Distribution of time and effort (mean points) dedicated to benefit-risk assessment of medicines.

On average, regulators allocate more time and effort in assessing favourable effects and their uncertainties than unfavourable effects and the uncertainties around them (EMA 2011a). Moreover, nearly all assessors start assessments by analysing the upper left quadrant because according to them if there is no favourable effect, there is no need to assess the rest. After starting the assessment, there was no preferred order of assessment to follow with regards to the remaining components, probably because assessors do not conceptualize the assessment process in their minds as reflected in the above diagrams. In addition, although there is a good convergence in defining benefits as "good things" (or favourable effects), there is lack of convergence in the definition, perception and interpretation of risks. The definitions previously given for explaining what is risk fall across the four quadrants,

although most of them belongs to the "bad things" (or unfavourable effects) quadrant, others can be framed in the remaining quadrants: "uncertainty of bad things" (e.g., frequency of side effect), "uncertainty of realizing a good thing" and "good things" (i.e., risk as the lack of benefit) (EMA 2011a).

In view of these results, the EMA updated the "Guidance for the CHMP Day 80 Assessment Report" in order to reflect the four-fold model, defining benefits as "favourable effects" and risks as "unfavourable effects" (EMA 2015a). This guidance asks for the description of favourable effects as well as the uncertainties around favourable effects; and for the description of unfavourable effects as well as the uncertainties around unfavourable effects. Definitions for favourable and unfavourable effects are provided below (EMA 2015a):

- Favourable effect: "Any beneficial effect for the target population (often referred to as "benefit" or "clinical benefit") that is associated with the product. These commonly include improvements in clinical efficacy but are not limited to efficacy (for example, a reduction in toxicity could also be a favourable effect)";

- Unfavourable effect: "Typically, this would include any detrimental effects (often referred to as "risks", "harms" or "hazards" both known and unknown) that can be attributed to the product or that are otherwise of concern for their undesirable effect on patients' health, public health; or the environment". "Unfavourable effects are not necessarily limited to safety endpoints (e.g. unfavourable effects may also be loss of efficacy on some important efficacy endpoints or other undesirable effect)".

In practice, benefit may comprise "the combined expected values of several possible favourable clinical and health outcomes"; while risk usually stands for "the combined probabilities and magnitude of several potential harms, or negative clinical and health outcomes" (Ma et al. 2016).

Overall, as illustrated by the study promoted by the EMA, assessors from European regulatory authorities seem to find it more difficult to define, interpret and assess risks and the uncertainties around risks than benefits and their uncertainties. Such findings highlight the need of further research in order to improve the assessment of benefit-risk ratios of medicines, particularly in the side of risk assessment.

I.1.3. UNCERTAINTY AROUND BENEFITS AND RISKS VERSUS MARKET ACCESS

Clinical data collected from clinical trial phases I to IV is typically the core information used by regulatory authorities to assess benefit-risk of medicines, particularly when manufactures are seeking a marketing authorization (Leong, Salek & Walker 2015). As noted before, benefit-risk assessments are made under circumstances of uncertainty, reflecting the uncertainties around efficacy and safety parameters (Eichler et al. 2008). The discussion on clinical efficacy and/or clinical safety are often the most important parts of assessment reports (EMA 2015a).

I.1.3.1. Clinical efficacy

Clinical trials, which are used to feed pre-marketing assessment reports, are primarily designed and statistically powered to provide reliable and robust conclusions on the clinical efficacy of medicines through the investigation of pre-defined endpoints (Leong, Salek & Walker 2015).

Usually, the primary endpoint of a clinical trial is an efficacy variable, as the primary objective of most clinical trials is to provide evidence on efficacy of treatments (ICH 1998). The primary endpoint should provide the most clinically relevant and convincing evidence related to the primary objective of a clinical trial. It should be well defined and capable of providing a valid and reliable measure of clinically relevant and important treatment benefit in the patient population intended for the treatment (ICH 1998), (Leong, Salek & Walker 2015). Those properties are assessed through the evaluation of content validity, which is "the extent to which an instrument measures the important aspects of concepts most significant and relevant to the patient's condition and its treatment" (Patrick et al. 2011). Further characteristics of the primary endpoint include sensitivity to the effects of the treatment, as well as being readily measurable and interpretable (Fleming & Powers 2012).

The selection of the primary endpoint in a clinical trial is based on the fact that effects on such endpoint provide reliable evidence about whether the intervention provides clinically meaningful benefit to the patient (Fleming & Powers 2012). The primary outcome measure should be a "clinical event relevant to the patient" (Fleming & DeMets 1996), or an endpoint that "measures directly how a patient feels, functions or survives" (Temple 1995), (Fleming & Powers 2012).

However, when direct assessment of the clinical benefit to the patient through observing actual clinical efficacy is impractical (e.g. time horizon), surrogate endpoints may be used as outcome measures to predict clinical benefit (ICH 1998), (Leong, Salek & Walker 2015).

A surrogate endpoint is a biomarker "used as a substitute for a clinically meaningful endpoint", and therefore "changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint" (Temple 1995). A biomarker is defined as "a characteristic that is objectively measured and evaluated as an indication of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (Biomarkers Definitions Working Group 2001).

There are several examples of surrogate endpoints accepted by regulatory authorities as good predictors of clinical benefit in the patient population intended for treatment: low-density lipoprotein (LDL) cholesterol in heart disease, blood pressure in hypertension, glycated haemoglobin (HbA1c) in diabetes mellitus, or tumour imaging and progression-free survival (PFS) in oncology (Lathia et al. 2009), (Liberti et al. 2015). However, there are also few examples of unsuccessful surrogate endpoints that in theory would have biologic plausibility, but did not result in clinical benefit for the patient; for example, high-density lipoprotein (HDL)/LDL cholesterol for oestrogen therapy in the prevention of cardiovascular disease (CVD) (oestrogen reduces cholesterol and epidemiologic trials show increased CVD after menopause, but there is no evidence of improved cardiovascular outcome with therapy, despite cholesterol reduction) (Lathia et al. 2009). The use of duly validated surrogate endpoints is necessary to predict clinical benefit. In addition, positive effects on surrogate endpoints do not necessarily result in clinical benefits for the patient (ICH 1998).

The selection of primary endpoints that directly translate to unequivocal benefits for the patient is of utmost importance for benefit-risk assessment (Ma et al. 2016). Of recall, different stakeholders may have different views on what constitutes benefit. Thus, the endpoints of a well-designed clinical trial may not produce a measurable meaningful benefit for the patient population (Leong, Salek & Walker 2015). The authors of a study reviewing marketing authorization applications submitted to the FDA between 2000 and 2012 for new molecular entities found that 151 (out of 302; 50%) applications were unsuccessful upon the first submission; of those, 20 (13%) were not approved because the study end points were poorly selected and therefore failed to adequately reflect a clinically meaningful effect (Sacks

et al. 2014). Thus, although a clinical trial may show clinical efficacy on a given endpoint, such evidence do not necessarily mean benefit for the patient.

Further, although clinical investigation is carried out under controlled conditions and statistically powered to detect differences between interventions on efficacy parameters, there are still some issues (e.g. gross experimental error, systematic error and bias, random error) that can lead to uncertainties around results on clinical efficacy endpoints (Eichler et al. 2008), (Mills et al. 2015). The results of a clinical trial may be subject to random error, which can be diminished but not eliminated, i.e. type I (wrong conclusion that one treatment has greater efficacy than another but, in fact, it has not) and type II errors (wrong conclusion that there is no difference between treatment efficacy but, in fact, there is) (Rothman 2010), (Akobeng 2016), (Bratton et al. 2016), (Bhatt & Mehta 2016). Further, gross experimental error (e.g. the use of non-validated surrogate endpoints) and bias (e.g. selection bias, allocation bias) are still common, and may result in challenging assessments to regulatory authorities (Fleming & Powers 2012), (Weintraub, Lüscher & Pocock 2015), (Yu et al. 2015), (Savović et al. 2012), (Clark, Fairhurst & Torgerson 2016), (Paludan-Müller, Teindl Laursen & Hróbjartsson 2016), (Eichler et al. 2016). Such limitations and uncertainties have to be taken into account and should be pointed out for the assessment of benefits and weighed for the assessment of benefit-risk balances (EMA 2015a).

I.1.3.2. Clinical safety

The pre-marketing assessment of risk or harm is mainly based on safety data collected from clinical studies that are typically designed and powered to prove clinical efficacy, not safety (Leong, Salek & Walker 2015). As such, uncertainty about safety may be even greater as compared to efficacy (Eichler et al. 2008). This may contribute to increased subjectivity in both the perception and conclusion on risks, as well as in the translation of safety information into objective outcomes (Slovic et al. 2004), (Leong, Salek & Walker 2015).

The outcome of regulatory decisions about benefit-risk assessments depends upon the level of uncertainty that regulatory authorities are willing to accept. This threshold of acceptable uncertainty is not static and varies across different therapeutic indications. Regulatory authorities are usually susceptible to accept more uncertainties about the

benefit-risk assessment of medicines for life-threatening diseases, namely those of high unmet medical need such as cancer, than for less severe conditions (Leyens et al. 2015).

In order to ensure early market access, regulatory authorities have been using at least one of the three main strategies: biomarkers and/or surrogate endpoints (discussed above); results from interim analyses (of value if able to anticipate unexpectedly large treatment effects when properly planned); and reduced-size safety databases (Eichler et al. 2008).

The use of a reduced-size safety database entails the question around how much safety data is needed for regulatory authorities to allow a medicine onto the market. In this context, it is useful to distinguish between predictable and unpredictable adverse drug reactions.

Clinical safety issues can be foreseen based on signals, such as the chemical structure and physicochemical properties of the active substance, primary and secondary pharmacology, metabolism, and findings from preclinical and clinical studies or from post-marketing experience with medicines in the same or similar class (Eichler et al. 2008). It is not quite understandable if regulatory authorities do not seek for full understanding of the safety profile and its implication for public health when safety signals arise in pre-marketing studies. Rofecoxib (Vioxx®; Merck) provides an example of a medicine for which there were pre-marketing signals indicating a potential for an increased cardiovascular risk with the medicine (Jüni et al. 2004). Regulatory authorities were criticized because of approving a medicine for which there was preliminary evidence of unacceptable risks (The Medicines in Europe Forum 2007).

Then, there are idiosyncratic adverse drug reactions, which are unpredictable based on the pharmacology of the drug (Rawlins & Thompson 1977), (Uetrecht & Naisbitt 2013). Although this type of reactions is usually rare, it often causes serious harm for the patient. These adverse drug reactions are usually not observed during pre-marketing clinical trials, and therefore there are few examples of medicines withdrawn from the market due to serious idiosyncratic reactions (Uetrecht & Naisbitt 2013). Valdecoxib (Bextra®; Pfizer), for example, was withdrawn from the market after several spontaneous reports of unpredictable serious skin reactions have been received by regulatory authorities (EMA 2005a), (EMA 2005b). For this type of adverse reactions, even the most rigorous regulatory criterions and a delay in marketing approval would not guarantee a favourable benefit-risk balance in post-marketing (Eichler et al. 2008).

As illustrated in Figure I. 4, an unrealistically large minimum number of patients would need to be enrolled in a clinical trial in order to demonstrate absence of risk for adverse drug reactions that are rare in nature. As an example, a randomized controlled trial including 7,000 patients in two treatment groups (i.e. a very large trial as compared to common standards) would have <50% power to detect an increase of 1 in 500 incidence of an adverse drug reaction that has a background incidence of 1 in 1,000 patients (one-sided test alpha = 5%). If that adverse drug reaction was fatal, an incidence of 1 in 5,000 (one-fifth of the background incidence) would probably shift the benefit-risk balance towards negative, and the medicine would not receive market approval (200 people would die for each one million receiving the medicine) (Eichler et al. 2008). The trial with 7,000 patients would not have enough power to detect such increase. A randomized controlled trial would need to include nearly one million patients to detect this adverse drug reaction with 90% power, which is unfeasible in practice (Rawlins 2004), (Schultz 2007), (Eichler et al. 2008).

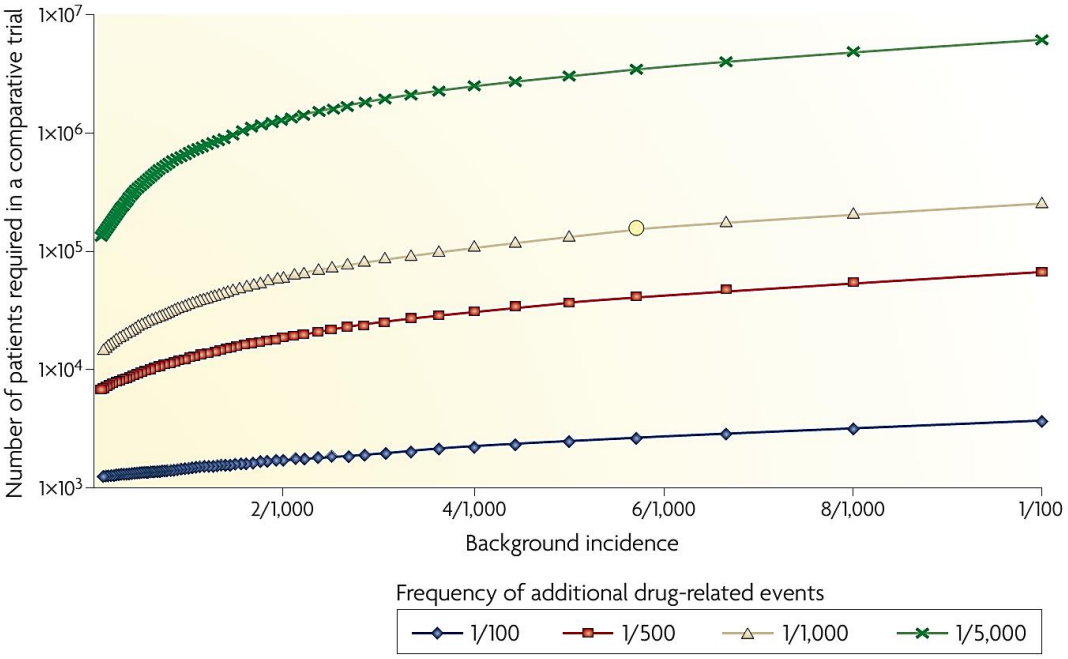


Figure I. 4 – Detection of rare adverse drug reactions in clinical trials.

This figure shows that over 160,000 patients would need to be included in a clinical trial to detect a 1/1,000 incidence of an adverse drug reaction, given a background incidence rate of 6 per 1,000 (large round symbol in graph; one-sided test, alpha = 5%, power = 80%).
Source: (Eichler et al. 2008).

Based on the example provided above, rare adverse drug reactions will continue to be identified only after large exposition during post-marketing; small increases of incidence of relatively common events will continue to be difficult to detect in pre-marketing clinical trials; and the safety database of a clinical trial would have to be expanded by a very large magnitude so that a pre-marketing study would have enough power to detect rare adverse

drug reactions. Thus, one may conclude that medicines will continue to be withdrawn from the market due to rare and serious adverse drug reactions that are only detectable after post-marketing (Eichler et al. 2008).

I.1.4. EARLY PATIENT ACCESS TO MEDICINES: TOOLS TO TACKLE BENEFIT-RISK UNCERTAINTIES

At the point of product approvals, there is limited information on potential risks. This is mitigated by post-marketing risk management plans and pharmacovigilance activities to monitor the safe use of the product, so as not to impede the timely access of potentially useful medicines (Leong, Salek & Walker 2015).

I.1.4.1. Conditional approvals

Both in the EU and the USA regulatory approaches are in place to facilitate early patient access to medicines that are of major public interest and fill an unmet medical need[1]. These tools can be either based on early, interactive and continuous dialogue or on risk-based marketing authorization approaches (Leyens et al. 2015), (EMA 2016b), (FDA 2015). Table I. 1 presents the regulatory tools that can be used to facilitate early patient access. Early and continuous dialogue and faster evaluation tools are not further discussed.

Table I. 1 – Regulatory tools to foster patient's early access to new medicines.

Characteristic	EMA	FDA
Early and continuous dialogue	Adaptive pathways PRIME scheme	Fast track Breakthrough therapy designation
Faster evaluation	Accelerated assessment	Fast track Priority review
Less evidence	Conditional Approval Exceptional circumstances	Accelerated Assessment

Adapted from (Leyens et al. 2015).

Risk-based approaches entails programmes that accept higher levels of uncertainty for medicines showing encouraging early efficacy results and acceptable safety, with sharing of risks between regulatory authorities, the public and the marketing authorization holder (Leyens et al. 2015). These approaches are called conditional approvals in the EU and accelerated assessments in the USA (EMA 2016b), (FDA 2015).

[1] Unmet medical needs mean a condition for which there exists no satisfactory method of diagnosis, prevention or treatment in the EU or, even if such a method exists, in relation to which the medicinal product concerned will be of major therapeutic advantage to those affected (EMA 2016b).

Since early approvals are supported by less comprehensive data than typically required, holders of medicines with conditional marketing authorizations are legally obliged to provide further evidence that confirms the initial benefit-risk evaluation, such as data from ongoing or new clinical studies, or pharmacovigilance reports (EMA 2016b).

A study found that, between 2007 and 2015, half of conditional marketing authorizations granted in the EU was supported by data from phase II clinical trials; and 43% of trials were single-arm uncontrolled studies (Leyens et al. 2015). Surrogate endpoints were used to support conditional approvals for all medicines authorized in the USA and 64% in the EU (Leyens et al. 2015). In terms of post-authorization requirements, while the EMA required more often (41%) the conclusion of ongoing trials and reporting of final results, the FDA asked more frequently (46%) for new confirmatory randomized controlled trials (Leyens et al. 2015).

Overall, these approaches, which have been used to assure that patients have timely access to new medicines, have been well succeeded. The early approval of medicines for unmet medical needs under exceptional circumstances or conditional approval procedures has not been associated with higher risk of serious safety issues (safety alerts and safety-related withdrawals) emerging after market approval, as compared to other medicines approved under standard pathways (Arnardottir et al. 2011).

However, a study that reviewed 26 cases of medicines receiving conditional approvals from the EMA found that there were delays or discrepancies in the fulfilment of post-approval obligations for more than one third of the procedures (Banzi et al. 2015). Another study reviewing the characteristics of post-marketing studies attached as specific obligations to the license of conditionally authorized medicines in the EU found that most of these studies (76%) were completed, but half were completed with a substantial delay (Hoekman et al. 2016). Thus caution is recommended with regard to broadening the use of these regulatory tools to resolve uncertainties around benefits and risks of medicines during post-marketing (Hoekman et al. 2016).

I.1.4.2. Risk management system

Randomized clinical trials are typically considered the highest-quality evidence in traditional evidentiary hierarchies (Guyatt et al. 2006), (Brozek et al. 2009). However, they have limitations in capturing safety information (Hammad, Pinheiro & Neyarapally 2011), (Hammad et al. 2013). Limiting factors may include a relatively small number of subjects,

short length of exposure and follow-up, restricted populations (e.g. age, ethnicity), as well as concomitant morbidities and concomitant medication, and lack of statistical power to assess multiple outcomes (EMA 2014a). In addition, some safety data from pre-marketing clinical trials may be not generalizable to the real-world setting, impacting post-marketing benefit-risk assessment and regulatory decision making (van Staa et al. 2008).

Thus, the knowledge about the benefit-risk balance of medicines at the time of their approval is limited, mainly because of uncertainties around clinical safety. Despite some actual or potential risks are notorious during clinical development, many others will be identified and characterized only after marketing. Since the assessment of benefit-risk balances is an ongoing and dynamic process, regulatory authorities entail a life-cycle risk management approach in order to allow for risk identification and characterization, as well as for risk minimisation or mitigation whenever possible. The main goal is to ensure that the benefits of a given medicine exceed the risks in the target patient population (EMA 2014a).

To achieve the above goal, marketing authorization holders must establish post-marketing risk management systems (RMS), in accordance to the European legislation (European Union 2001), (European Union 2004), (EMA 2014a).

A RMS is defined as "a set of pharmacovigilance activities and interventions designed to identify, characterise, prevent or minimise risks relating to medicinal products including the assessment of the effectiveness of those activities and interventions" (European Union 2001), (EMA 2014a). The RMS is intended to ensure a continuous monitoring of safety (i.e. pharmacovigilance) data in order to determine "whether there are new risks or whether risks have changed or whether there are changes to the benefit-risk balance" (European Union 2001).

Furthermore, a detailed risk management plan (RMP), which is a detailed description of the RMS, must be submitted as part of marketing authorization application dossier (European Union 2001), (EMA 2014a). Each RMP is discussed and agreed between the marketing authorization holder and the regulatory authority upon the licensing of a medicine. Although the RMP is primarily focused on risks, the need for efficacy studies must be evaluated and, if needed, incorporated in the RMP as such information is important to assess risks into context. The RMP is continuously revised and updated during the lifetime of a medicine, as new information becomes available. The principles of risk management are illustrated by Figure I. 5.

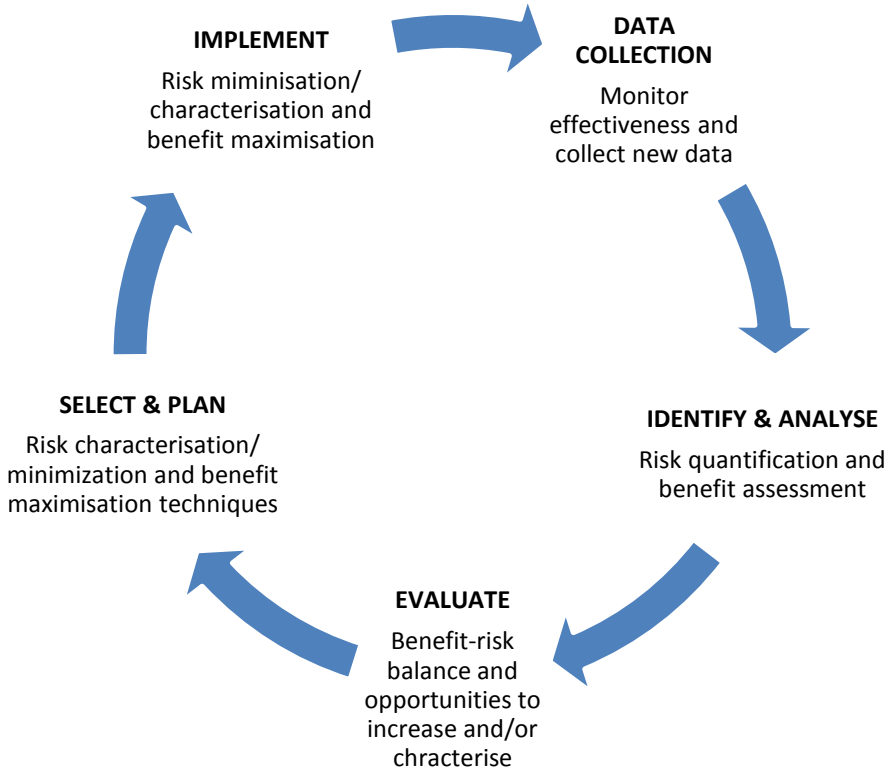


Figure I. 5 – The risk management cycle.

Source: (EMA 2014a).

The RMP should be designed in a way that addresses the safety concerns found for the medicine, including important identified risks, important potential risks, and important missing information. The plan comprises routine pharmacovigilance activities (e.g. collection, collation, assessment and reporting of spontaneous reports of suspected adverse reactions) and, if needed, additional pharmacovigilance activities that may be non-clinical studies, clinical trials or non- interventional studies (EMA 2014a). These additional studies may be voluntarily proposed by marketing authorization holders or imposed by regulatory authorities (for example, as condition to approve the marketing of a medicine), and includes the conduction of post-authorisation safety studies (PASS), and/or post-authorization efficacy studies (PAES) (European Union 2001).

PASS are aimed at identifying, characterising or quantifying a safety hazard, confirming the safety profile of a medicine or of measuring the effectiveness of risk management measures (EMA 2016c). PAES are usually required where concerns relating to some aspects of the efficacy of a medicine are identified and can be resolved only after marketing (European Union 2001).

Post-marketing studies, namely PASS, may encompass a variety of designs, i.e. different epidemiological methods (EMA 2016c):

- Intensive monitoring schemes
    - System of record collation in designated areas (e.g. hospital units or by specific healthcare professionals in community practice);
    - Implicate reviewing medical records or interviewing patients and/or healthcare professionals;
    - Allow to collect information about events, use of medicines (e.g. determine the potential for abuse), and specific subgroups of patients;
    - Limitations may include selection bias, small number of patients and high costs.

- Prescription event monitoring
    - Uses electronic prescription data or automated health insurance claims to identify patients;
    - Information about outcomes (e.g. adverse events) and others (e.g. characteristics of patients, use of medicine) is obtained through questionnaires sent to prescribing physicians or patients;
    - Limitations may include incomplete response from interviewees.

- Registries
    - System that uses observational methods to collect uniform data on given outcomes in a population with a particular disease, condition or exposure;
    - Disease registries or exposure registries, depending on the type of information primarily entered (diagnosis of diseases or prescription of a medicine, respectively);
    - Disease registries may help collect data on medicines exposure or risk factors for a clinical condition;
    - Disease registries may serve as base for case-control studies that compare exposure to medicines in cases from the registry with that in controls (selected from the same registry but without the condition or from other registries).

  - Exposure registries are focused on populations exposed to medicines of interest and intend to explore the effects of the exposition;
  - Exposure registries allow for following patients over time and collecting data on adverse events using standardised questionnaires (e.g. cohort study to determine event incidences).

- Observational studies
  - Cross-sectional studies (surveys);
  - Cohort studies;
  - Case-control studies, and nested case-control studies;
  - Other designs (self-controlled case-series, case-crossover, and case-time control).

- Clinical trials
  - Post-marketing clinical trials, comprising for example pharmacodynamic and pharmacokinetic assessments, or genetic testing, can be useful to evaluate mechanisms of adverse reactions and identify subgroups of patients that are at increased risk;
  - The magnitude of risk or benefit can be studies in special populations often excluded from pre-marketing studies (elderly, children, or patients with renal or hepatic disorder).

- Drug utilization studies
  - These studies aim to describe how a medicine is prescribed and used in daily clinical practice in large populations;
  - Allow for the characterization of patients and evaluate how given characteristics impact clinical, social and economic outcomes;
  - They can be used, for example, to estimate rates of adverse reactions, monitor the effect of regulatory actions, or audit actual clinical practice by comparing it with recommendations or guidelines.

The establishment of a formal RMS approach results from the acknowledgement that passive pharmacovigilance systems alone, i.e. spontaneous reporting of suspected adverse drug reactions, are insufficient to ensure safety of all medicines allowed onto the market (Eichler et al. 2008). Although spontaneous reporting is of unquestionable value for pharmacovigilance, having supported numerous safety alerts, suspensions and withdrawals of medicines over the years (Alves, Macedo & Batel-Marques 2013), (La Rochelle, Lexchin & Simonyan 2016), some inherent features, such as underreporting, the lack of a denominator and controls, may diminish the ability to distinguish signals from background noise (Hazell & Shakir 2006), (Grundmark et al. 2014). Lately, an increase in the use of other sources of evidence (i.e. data from epidemiological and clinical studies) to support regulatory decisions on drug safety has been noted (Paludetto, Olivier-Abbal & Montastruc 2012), (McNaughton, Huet & Shakir 2014).

A more proactive approach, which includes structured PASS and PAES, is foreseen by the current European legislation in order to strengthen pharmacovigilance, to promote and protect public health by reducing burden of adverse drug reaction and optimising use of medicines (European Union 2001).

The introduction of RMP approaches have become a cornerstone in pharmacovigilance to support a proactive attitude for acquiring knowledge about safety profiles of marketed medicines. Nevertheless, the added value of this strategy should be addressed in future research, as evidence on its effectiveness is scarce.

A study, evaluating the evolution of safety concerns listed in the RMP of 48 medicines intended for chronic use, found that 20% of the pre-marketing uncertainties were resolved 5 years after approval, but new uncertainties had been included in the RMP at a similar rate (Vermeer et al. 2014). Further, there is a need to raise awareness among PASS stakeholders (i.e. regulatory authorities and sponsors) to increase the availability of protocols and theirs assessments, as well as to design more thoughtful studies that apply proper epidemiologic methods, have an adequate analytic plan and use right data sources (Engel et al. 2016).

Facing the challenge of enabling early access to medicines, regulatory authorities have been granting marketing authorizations for medicines with positive benefit-risk balances, but for which some uncertainties could remain. In a particular risk-averse environment, regulatory authorities found ways (as those outlined before) of continue approving medicines on a timely manner by conducting repetitive benefit-risk assessments over theirs life-cycles, including post-marketing phases.

The focus of regulation should aim the maximization of gains for public health, and not only risk minimization. Few strategies were outlined by members of the EMA staff with the aim to "best align acceptance of risk and uncertainty by regulators with the best interests of public health (Eichler et al. 2013), (EMA 2013a):

1. Define ways to systematically include the patient view on the level of acceptable risk linked to a medicine, which may be different to regulators' assumptions;
2. Reflect on methodologies to combine value judgements, including patients' values, with interpretation of 'hard' data;
3. Develop the concept of 'tolerability of risk' thresholds for medicines evaluation, recognising that zero risk situations do not exist in real-world conditions;
4. Take into account the shift in medicines regulation towards an emphasis on surveillance of safety and effectiveness in the real world. Development of robust tools to enable real-time knowledge generation, faster decision-making and opportunities for risk minimisation measures should reduce the perceived need for risk aversion at the initial stage of licensing;
5. Allow medicines regulators to factor 'opportunity costs' into standards for evaluation of the benefit-risk of medicines and in individual marketing-authorisation decisions".

As outlined in the fourth topic, the emphasis of drug regulation should address continuous monitoring of safety and effectiveness of medicines in real-world. This strategy entails challenging assessments of benefit-risk balances, as data from multiple sources need to be interpreted, valued and integrated from a clinical and statistical point of view to support regulatory decisions.

In this context, some researchers have claimed for new methodologies to integrate data coming from multiple sources (Hammad et al. 2013), (Alves, Batel-Marques & Macedo 2014). Further, a need for improvement in the clarity and transparency of benefit-risk assessments that support regulatory decisions has been asked given that this process is essentially a subjective qualitative weighing of the evidence that relies heavily on expert opinions (EMA 2007), (FDA 2013a). Thus, the introduction of structured frameworks that encompass quantitative methods for benefit-risk assessment may be of added value to better support more informed and science-based regulatory decisions (Guo et al. 2010), (EMA 2011b), (PROTECT 2013), (Mt-Isa et al. 2014), (Hallgreen et al. 2014), (Nixon et al. 2016), (Hughes et al. 2016).

I.2. NEED TO ENHANCE BENEFIT-RISK ASSESSMENTS

Over the past few years, several medicines were suspended or withdrawn from the market due to safety reasons. The list includes medicines indicated to treat diseases of considerable prevalence, such as troglitazone, cisapride, cerivastatin, rofecoxib or rosiglitazone (Mann & Andrews 2014). Regulatory authorities impose such measures when they come with the conclusion that the benefits of a medicine no longer outweigh its risks. These decisions have major impact in the society, affecting regulatory authorities, patients, healthcare professionals, and manufacturers (Onakpoya, Heneghan & Aronson 2016). In part due to the several cases of post-marketing withdrawals of medicines, the stakeholders involved in the field of drug regulation have focused their research on safety evaluation and improvement of processes for benefit-risk assessment. In addition, there are discrepancies in the patterns and inconsistencies across countries regarding the withdrawal of medicines from the market (Onakpoya, Heneghan & Aronson 2016), (Onakpoya, Heneghan & Aronson 2015). As an example, rosiglitazone was withdrawn from the European market due to cardiovascular safety issues, but it was allowed to continue being used in the USA (Mendes, Alves & Batel-Marques 2015). This case illustrates that two different regulatory authorities may reach contradictory conclusions despite analysing the same evidence.

According to Rawlins (1987), benefit-risk assessments comprehend formal, comparative and informal analyses (Rawlins 1987). While a formal analysis entails a science-based deductive process and quantitative comparisons of benefits and risks (expressed as numerical trade-offs), an informal analysis is an inductive process that involves personal judgement. As noted in the report issued by the Council for International Organizations of Medical Sciences (CIOMS) Working Group IV in 1998, probably the majority of benefit-risk assessments rely on a relatively informal basis, meaning that these assessments are based on the "fallibility of human judgement" (Rawlins 1987), (CIOMS Working Group IV 1998). By that time, the Working Group asked for further research on quantitative and semi-quantitative approaches that could help regulatory authorities and other stakeholders to actually quantify benefit-risk ratios of medicines, rather than continuing to depend on judgment alone (CIOMS Working Group IV 1998).

Although expert judgment is useful for valuing individual items of evidence, it has limitations when synthesizing multiple valuations is required (Edwards 1968), (Edwards et al. 1968) which is often the case in benefit-risk assessments (EMA 2007). Additionally, judgements may be biased due to heuristic approaches used to support deliberative

reasoning (Kahneman 2002), (Mellers & Locke 2007). In this context of complexity, namely regarding the process of synthesizing information, regulatory decisions are mostly implicit. Consequently, the communication of the reasons and rationale that support regulatory decisions is problematic (EMA 2007).

A review of the procedures followed by five European regulatory authorities found that benefit-risk assessments are made intuitively, the responsibility of an accountable senior assessor or of a team, as a result of extensive discussion (EMA 2007). Similarly, the assessments made within the FDA also rely heavily on clinical judgement after extensive analysis of evidence and discussion. Noteworthy, differences in clinical judgements among experts can lead to divergent individual opinions and conclusions on the benefit-risk assessment of a given medicine (FDA 2013a).

Overall, benefits and risks have been assessed in a holistically manner, based on intuitive approaches deprived of straightforward definitions of value structures and trade-offs (Pignatti et al. 2015). This process does not lead to explicit quantification of the risks and benefits and lacks objectivity, consistency, transparency, and reproducibility (Guo et al. 2010). Regulatory authorities need to use better methods for assessing benefit-risk balances and evolve from implicit to explicit decision-making (Eichler et al. 2012). This change requires the explicit description of all decision criteria, interpretation of data and also valuations, including weighing factors for treatment outcomes (Pignatti et al. 2015). The introduction of structured frameworks (that may encompass quantitative or semi-quantitative methodologies) has the potential to bring clarity to the process and help regulatory authorities to make more objective and evidence-based decisions (Guo et al. 2010), (Yuan, Levitan & Berlin 2011), (Mt-Isa et al. 2014).

I.2.1. PROJECTS FOR IMPROVING BENEFIT-RISK ASSESSMENTS

Regulatory authorities, pharmaceutical industries and academia have been working to develop and test approaches that can improve methodological and communication aspects of benefit-risk assessment (Pignatti et al. 2015). The main initiatives are illustrated by Figure I. 6.

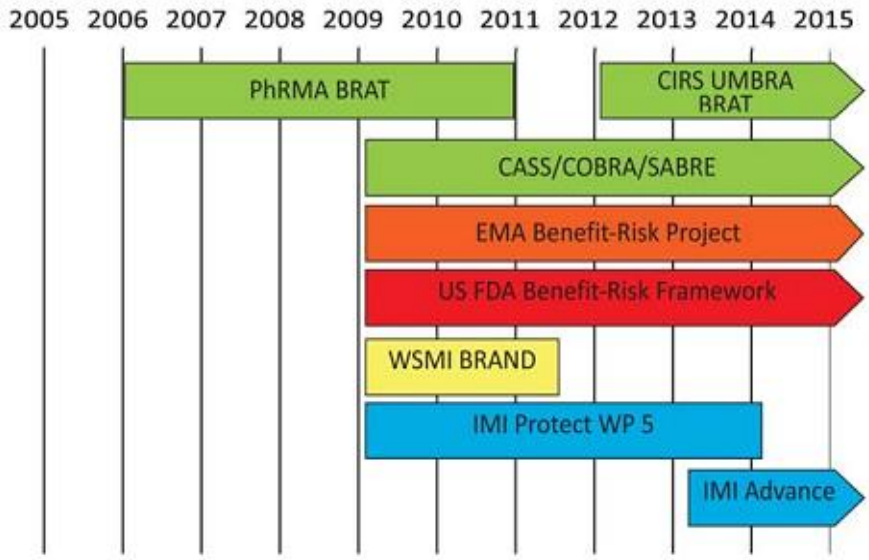


Figure I. 6 – Examples of benefit-risk initiatives.

Source: Pignatti et al. 2015
ADVANCE, Accelerated development of vaccine benefit-risk collaboration in Europe; CASS, Taskforce of representatives from Health Canada, Australia's Therapeutic Goods Administration, Swissmedic, and the Singapore Health Science Authority; CIRS, Centre for Innovation in Regulatory Science; COBRA, Consortium on Benefit-Risk Assessment; EMA, European Medicines Agency; FDA, USA Food and Drug Administration; IMI PROTECT WP5, Innovative Medicine Initiative Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium, work package 5; PhRMA BRAT, Pharmaceutical Research and Manufacturers of America Benefit-Risk Action Team; SABRE, Southeast Asia Benefit-Risk Evaluation; UMBRA, Unified Methodology for Benefit-Risk Assessment; WSMI BRAND, World Self Medication Industry Benefit-Risk Assessment for Non-prescription Drugs.

I.2.1.1. European initiatives

I.2.1.1.1. The EMA Benefit-Risk Methodology Project

The EMA Benefit-Risk Methodology Project was initiated in 2009 to explore development in methodologies for benefit-risk analysis, including the test of qualitative frameworks and quantitative approaches, with the aim of improving regulatory decision making about medicinal products (EMA 2009a).

The project was divided in five work packages. The objective of the second work package was to assess the applicability of existing tools and processes for benefit-risk assessment (EMA 2010a). A generic qualitative framework, called PrOACT-URL (problem, objectives, alternatives, consequences, trade-offs, uncertainty, risk tolerance, linked decisions) (Table I. 2), which follows the eight-stage general decision framework developed by (Hammond, Keeney & Raiffa 1999), was proposed for benefit-risk assessment decision-making by regulators (EMA 2010a).

Table I. 2 – EMA 8-Step PrOACT-URL.

Steps	Actions
1. Problem	• Determine the nature of the problem and its context. • Frame the problem.
2. Objectives	• Establish objectives that indicate the overall purposes to be achieved. • Identify criteria of favourable and unfavourable effects.
3. Alternatives	• Identify the options to be evaluated against the criteria.
4. Consequences	• Describe how the alternative performs for each of the criteria—that is, the magnitudes of all effects and their desirability or severity and the incidence of all effects.
5. Trade-offs	• Assess the balance between favourable and unfavourable effects.
6. Uncertainty	• Assess the uncertainty associated with the favourable and unfavourable effects. • Consider how the balance between favourable and unfavourable effects is affected by uncertainty.
7. Risk tolerance	• Judge the relative importance of the decision makers' risk attitude for this product and indicate how this affected the balance reported in step 5
8. Linked decisions	• Consider the consistency of this decision with similar past decisions, and assess whether taking this decision could affect future decisions

Source: (Walker et al. 2015).

The PrOACT-URL framework includes a tabular display, called "effects table", which lists important effects and their uncertainty (Table I. 3). The use of such tables can contribute to improve transparency of benefit-risk assessments and support communication between the stakeholders involved in the process (EMA 2014b). The "effects table" forms an integral part of the benefit-risk section of new drug applications assessment reports since 2015 (EMA 2015b).

Table I. 3 – Example of an Effects Table for vandetanib, based on the EPAR EMEA/H/C/002315.

	Effect	Short Description	Unit	Placebo	Vandetanib	Uncertainties/ Strength of evidence	References
Favourable	PFS (HR)	From randomization to progression or death (blinded independent review)	N/A	1	0.46 (95% CI: 0.31, 0.69)	Large effect in overall population. Consistent and significant effect on PFS but not OS (too early?)	See Discussion on Clinical Efficacy.
	PFS (median)	Weibull model	Mo	19.3	30.5	Only a very low number of patients with definite RET mutation negative status at baseline. Lower efficacy?	Single-arm study in RET negative patients post-approval.
	ORR	Proportion of complete or partial responders (>=30% decrease unidimensional) RECIST	%	13	45	No clear effect on PRO/QoL (missing data)	See Discussion on Clinical Efficacy.
Unfavourable	Diarrhoea Grade 3-4	Increase of ≥7 stools per day over baseline; incontinence; Life-threatening	%	2.0	10.8	Duration of follow up in the pivotal study is short vs. the need for long duration of treatment.	Risk of dehydration and renal/cardiac risks (see SmPC 4.4)
	QTc related events Grade 3-4	QTc >0.50 second; life threatening; Torsade de pointes	%	1.0	13.4	Risk of developing further major cardiac SAEs including Torsade de pointe?	Restrict to symptomatic and aggressive disease (see SmPC 4.1).
	Infections Grade 3-4	IV antibiotic, antifungal, or antiviral intervention indicated; Life-threatening	%	36.4	49.8		Explore lower dose (see Table 20. Summary of the RMP)

HR, hazard ratio; IV, intravenous; Mo, months; ORR, objective response rate; OS, overall survival; PFS, progression-free survival; PRO, patient reported outcome; QoL, quality of life; RET, rearranged during transfection (gene); RECIST, Response Evaluation Criteria In Solid Tumours; RMP, risk management plan; SAE, serious adverse event; SmPC, summary of product characteristics.
Source: (EMA 2014b).

Three other qualitative approaches, which were not fully assessed because of being under development at the time, were described, namely the PhRMA BRAT (Pharmaceutical Research and Manufacturers of America Benefit-Risk Action Team), the 7-step framework from the CIRS (Centre for Innovation in Regulatory Science), and the FDA benefit-risk framework (EMA 2010a). Further, 18 quantitative approaches were reviewed (EMA 2010a).

One of the main findings reached by the EMA was that the application of a quantitative method or approach requires a qualitative framework within which the model

can be effectively developed. Indeed, a qualitative approach may suffice for simpler benefit-risk decisions (EMA 2010a).

Further, combination of approaches could prove useful in situations characterized by several contributions, namely the magnitude of favourable effects, the seriousness of unfavourable effects, uncertainty about the effects, transitions in health states and the time spent in each state, and trade-offs between effects. In conclusion, the use of structured processes, both qualitative and quantitative, was thought to improve transparency, communicability, audibility, quality and speed of decision making (EMA 2010a).

In addition to the promotion of the current framework, i.e. the PrOACT-URL, the EMA has explored opportunities for implementing additional tools, including methods described in the Innovative Medicines Initiative PROTECT Project (Pignatti et al. 2015).

I.2.1.1.2. The IMI PROTECT Project

The Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT) is a multi-national consortium of 34 public and private partners, including regulatory authorities, pharmaceutical companies and academics, coordinated by the EMA. The main goal is to address limitations of methods used in pharmacoepidemiology and pharmacovigilance (PROTECT 2011).

One of the working programs, WP5, was specifically designed to develop methods for use in benefit-risk assessment, in particular to (PROTECT 2011):

- Identify, characterize and test methods of collating data on benefits and risks from various data sources, parameters and strengths of evidence, and of integrating them with decision-criteria and formal assessment of values of patients, healthcare providers, regulators, the pharmaceutical industry and in benefit-risk assessment;
- Identify, test and compare modelling approaches that would allow continuous benefit-risk risk-modelling along the lifecycle of the product, and support decision-making;
- Develop methods of graphical expression of the benefits and risks of the medicinal products for use by patients, healthcare providers, the pharmaceutical industry and regulators along the lifecycle of the product.

The PROTECT Benefit-Risk group has developed recommendations for benefit-risk decision processes and supporting tools, which were organized around the five stages of a generic benefit-risk assessment roadmap (Figure I. 7; Table I. 4).

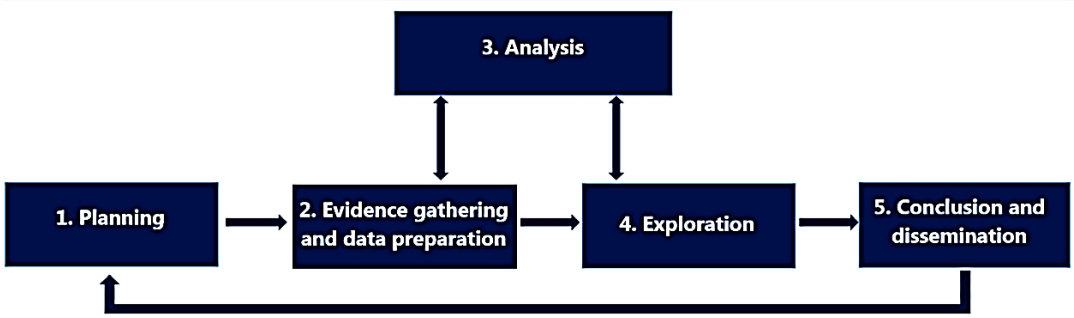


Figure I. 7 – Five stages of a generic benefit-risk assessment.

Table I. 4 – Description of the five stages followed in a generic benefit-risk assessment.

Stage	Description
Planning	• A descriptive framework, such as BRAT or PrOACT-URL, should be used to structure each benefit-risk assessment; • A set of benefits and risks should be chosen that covers the full range of treatment effects, and represented visually using a tree diagram to indicate the hierarchy; • A table template ("effects table" or "source table") should be prepared to represent the data that are required to be collected.
Evidence gathering and data preparation	• Assessors should review all available evidence and select data that are sufficient to and appropriate for the decision problem; • The table template must be completed, highlighting where data are available or missing (for example by colour-coding missing data); • The tree diagram and table produced initially may need to be revised in the light of available data.
Analysis	• The analysis should be appropriate to the complexity of the task; • Simple descriptive methods may suffice for routine benefit-risk assessments, while quantitative decision models can provide additional clarity for more complex problems; • When a quantitative approach is used, value preferences and the magnitudes of benefits and risks (by criteria and overall) should be presented by suitable bar graphs, dot plots or line graphs to promote accurate point reading, comparisons and judgment of trade-offs among alternatives.
Exploration	• This stage assesses the robustness and sensitivity of the results; • Quantitative decision models facilitate de execution and communication of sensitivity analyses by setting out the impacts of effects uncertainty and preference uncertainty on the results; • Preferred visualisation techniques include distribution plots, line graphs, forest plots or tornado diagrams to provide comprehensive overview of the benefit-risk analysis allowing better informed decisions.
Conclusion and Dissemination	• In this stage, a conclusion is reached after considering all the information from previous stages; • Adoption of a formal structure for benefit-risk assessment allows for an effective way to improve the overall transparency and communication of the process and facilitate robust decision making.

BRAT, Benefit-Risk Action Team; PrOACT-URL, Problem, Objectives, Alternatives, Consequences, Trade-offs, Uncertainty, Risk and Linked decisions.
Source: (Hughes et al. 2016)

The project from the PROTECT Benefit-Risk group comprised two core areas of research, namely "Benefit-Risk Assessment Methodologies" and "Benefit-Risk Assessment Visual Representations". The latter is not further addressed.

A systematic review of the literature was performed by researchers of the PROTECT group to identify, appraise and classify available benefit-risk methodologies with the aim to facilitate and inform their future use. The authors identified 49 methodologies, which were classified into four categories (Figure I. 8) (Mt-Isa et al. 2014):

- Benefit-risk assessment frameworks;
- Metric indices;
- Estimation techniques; and
- Utility survey techniques.

Each category was described by the authors as follows (Mt-Isa et al. 2014):

- Frameworks are structured stepwise methodologies to perform a task, which can be descriptive (i.e. provide qualitative stepwise instructions) or quantitative (i.e. provide explicit methods for balancing benefits and risks);
- Metric indices are systems of measurement that consist of those that provide thresholds of benefits or risks (i.e. handle either benefit or risk, but not both) or those that essentially weigh the benefits and risks, namely health indices (i.e. validated and standardized quality-of-life indicators) and trade-off indices (i.e. methods that integrate benefits and risks into a single metric representing the value of the trade-off for direct interpretation of whether a treatment option is favourable or unfavourable);
- Estimation techniques include generic statistical techniques, which are not unique to benefit–risk assessment, but are readily applicable in combination with other benefit–risk methods. They can be useful in synthesizing evidence from multiple sources and in handling statistical uncertainties in the decision model;
- Utility survey techniques include methods to elicit and collect utilities and value preferences of various outcomes. These techniques are not specifically benefit-risk assessment methods, but are used in combination with other benefit–risk methods to achieve more robust utility values and to increase the transparency of the decisions.

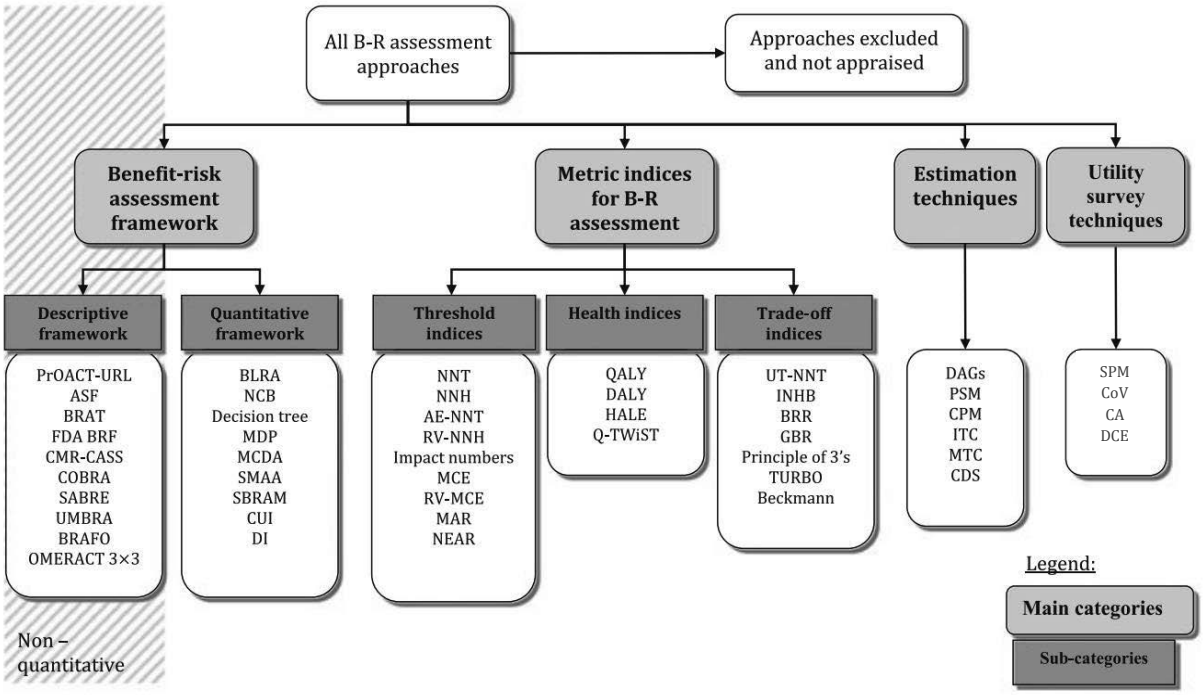


Figure I. 8 – Classification of methodologies used for benefit-risk assessment.

AE-NNT, adverse event adjusted number needed to treat; ASF, Ashby and Smith Framework; BLRA, benefit-less–risk analysis; BRAFO, Benefit–Risk Analysis for Foods; BRAT, Benefit-Risk Action Team; BRR, benefit–risk ratio; CA, conjoint analysis; CDS, cross-design synthesis; CMR-CASS, Centre for Medicines Research Health Canada, Australia's Therapeutic Goods Administration, Swissmedic, and the Singapore Health Science Authority; COBRA, Consortium on Benefit-Risk Assessment; CPM, confidence profile method; CUI/DI, clinical utility index/desirability index; CoV, contingent valuation; DAG, directed acyclic graphs; DALY, disability-adjusted life years; DCE, discrete choice experiment; FDA BRF, USA Food and Drug Administration Benefit-Risk Framework; GBR, global benefit–risk; HALE, health-adjusted life years; INHB, incremental net health benefit; ITC, indirect treatment comparison; MAR, maximum acceptable risk; MCDA, multi-criteria decision analysis; MCE, Minimum clinical efficacy; MDP, Markov decision process; MTC, mixed treatment comparison; NCB, net clinical benefit; NEAR, net efficacy adjusted for risk; NNH, number needed to harm; NNT, number needed to treat; OMERACT 3x3, Outcome measures in rheumatology 3 × 3; PrROACT-URL, Problem, Objectives, Alternatives, Consequences, Trade-offs, Uncertainty, Risk and Linked decisions; PSM, probabilistic simulation method; QALY, quality-adjusted life year; Q-TWiST, quality-adjusted time without symptoms and toxicity; RV-MCE, relative value-adjusted minimum clinical efficacy; RV-NNH, relative value-adjusted number needed to (treat to) harm; SABRE, Southeast Asia Benefit-Risk Evaluation; SBRAM, Sarac's benefit–risk assessment; SMAA, stochastic multi-criteria acceptability analysis; SPM, stated preference method; TURBO, transparent uniform risk–benefit overview; UMBRA, Unified Methodology for Benefit-Risk Assessment; UT-NNT, utility-adjusted and time-adjusted number needed to treat.
Source: (Mt-Isa et al. 2014).

After reviewing the 49 methodologies, the authors concluded that there is not a "one-size-fits-all" method, and a combination of methods may be needed for each benefit-risk assessment (Mt-Isa et al. 2014). After taking into account the limitations and strengths specific to each methodology, the researchers recommended 13 methodologies for further examination in real-life benefit–risk assessment of medicines: two descriptive frameworks (PrOACT-URL and BRAT), two quantitative frameworks (MCDA and SMAA), two threshold indices (NNT [and NNH], and impact numbers) two health indices (QALY and Q-TWiST), two trade-off indices (INHB and BRR), two estimation techniques (PSM and MTC) and one utility survey technique (DCE) (Mt-Isa et al. 2014).

The recommended methodologies were further tested by the PROTECT group using eight case studies that were selected based on real-world scenarios involving medicines with

marginal benefit–risk balance and presenting various practical challenges to test candidate methods, namely efalizumab, natalizumab, rimonabant, rosiglitazone, telithromycin and warfarin (Hughes et al. 2016).

I.3. NUMBER NEEDED TO TREAT

I.3.1. BRIEF HISTORICAL BACKGROUND AND OTHER MEASURES OF TREATMENT EFFECT

There are several ways of expressing resulting effects of clinical interventions. However the results are not always easily transposed to clinical decision making. As an example, the results are frequently expressed in terms of risk, which is the probability of a certain event occur in a group. Although separate risks are useful when assessing two groups individually, comparative results (i.e. the outcome in one group relative to the outcome in the other group) are more informative to both clinicians and patients (McQuay & Moore 1997).

There are various treatment effect measures allowing to compare risks between groups. Table I. 5 provides an example of a 2x2 table that is used to illustrate those comparisons.

Table I. 5 – Example of a 2x2 table for assessing risk of an event in two groups.

	With event	Without event	Total
Treatment group	*a*	*b*	*a + b = N1*
Control group	*c*	*d*	*c + d = N2*
Total	*a + c = M1*	*b + d = M2*	*T = a + b + c + d*

Note: The event rate (or the probability of the event) in treatment group and control groups is given by P1=a/N1 and P2=c/N2.

I.3.1.1. Relative risk and relative risk reduction

The relative risk (RR), that is the ratio between risks, is used to compare risks between groups. The risk of having an event is P1=a/N1 in the treatment group and P2=c/N2 in the control group. Thus, the RR is given by the ratio between the risk of the event in the treatment group and the risk of the event in the control, i.e. RR = P1/P2 (Table I. 5).

The relative risk reduction (RRR) is given by the difference in event rates between control group and treatment group, divided by the event rate in control group, that is RRR = (P2-P1)/P2 (Table I. 5). The RRR can also be calculated as 1 – RR.

Table I. 6 provides the results of a hypothetical parallel group clinical trial with a fixed follow-up time, in which patients were randomly allocated to receive active treatment or

placebo (control) to prevent stroke. The results are displayed according to the degree of hypertension presented by patients at the baseline.

Table I. 6 – Hypothetical clinical trial results.

	Stroke	No stroke	Event rate	RR	RRR	ARD	NNT
Moderate hypertension at baseline							
Treatment	1,800	13,200	0.12 (P1)	0.6	0.4	0.08	13
Control	3,000	12,000	0.20 (P2)				
Mild hypertension at baseline							
Treatment	135	14,865	0.009 (P1)	0.6	0.4	0.006	167
Control	225	14,775	0.015 (P2)				

ARD, absolute risk difference (ARD=P2-P1); RR, relative risk (RR=P1/P2); RRR, relative risk reduction (RRR=[P2-P1]/P2); NNT, number needed to treat (NNT=1/ARD).

As illustrated in Table I. 6, the rate of stroke in patients with moderate hypertension is approximately 13 times higher than in those with mild hypertension. However, the RR (=0.6) (and the RRR, 0.4) was the same in both populations (Table I. 6). These results are interpreted as follows: the patients receiving treatment had 0.6 times the risk of stroke compared to patients receiving placebo (or the patients receiving treatment had a 40% reduction in risk of stroke compared to those receiving placebo).

Thus, the RR (and the RRR) has the disadvantage that a given value is the same whether the risk with treatment decreases from 0.20 to 0.12, from 0.015 to 0.009, and so forth. Since the RR (and the RRR) does not reflect the magnitude of the risk without therapy, it is difficult to discriminate between small and large treatment effects.

I.3.1.2. Odds ratio

Odds are the ratio of the probability of an event occurring in a group, divided by the probability of that event not occurring in that group. According to Table I. 5, the probability of the event occur in the treatment group and in control group is given by P1 (=a/[a+b]) and P2 (=c/[c+d]), respectively. Thus, the odds in the treatment group and in control group is P1/(1-P1) and P2/(1-P2), respectively.

The odds ratio (OR) expresses the odds of a patient in treatment group having an event compared to a patient in control group, i.e. OR = [P1/(1-P1)] / [P2/(1-P2)]. This formula can be derived and presented as OR = (a x d)/(b x c). Thus, an OR of five (that is, five to one) mean that five people will experience the event for every one that does not (a

risk of five out of six or 83%). An OR of 0.5 seems less intuitive: 0.5 people will experience the event for every one that does not – this translates to one event for every two non-events (a risk of one in three or 33%) (Davies, Crombie & Tavakoli 1998).

Using the example provided in Table I. 6, the OR would be estimated at 0.545 ([0.12/(1-0.12)]/[0.20/(1-0.20)]) for the population with moderate hypertension and 0.596 ([0.009/(1-0.009)]/[0.015/(1-0.015)]) for the population with mild hypertension. Thus, the OR is almost similar across the two populations, despite the rates of events are very different.

The OR is used as an approximation of the RR in case-control studies, but it can also be used as a measure of treatment effect in randomized trials. The RR can be calculated only if it is possible to estimate probabilities of an outcome in each group, which is not possible in case-control studies, where cases and controls are randomly selected. Further, the OR is often a statistic of choice in meta-analyses, given that it is more stable than other measures of treatment effect when applied across studies with various incidence rates.

Noteworthy, the OR is close to the RR if probabilities of the outcome are small: OR = [P1/(1-P1)] / [P2/(1-P2)] = (P1/P2) x [(1-P2) / (1-P1)] = RR x [(1-P2) / (1-P1)] (Davies, Crombie & Tavakoli 1998), (Zhang & Yu 1998). However, the more frequent the outcome becomes, the more the OR will overestimate the RR when it is more than 1 or underestimate the RR when it is less than 1, as illustrated in Figure I. 9 (Zhang & Yu 1998). Thus, caution is needed when interpreting results of OR as thought to be a RR because it could be perceived as an effect size bigger than is actually the case (Davies, Crombie & Tavakoli 1998).

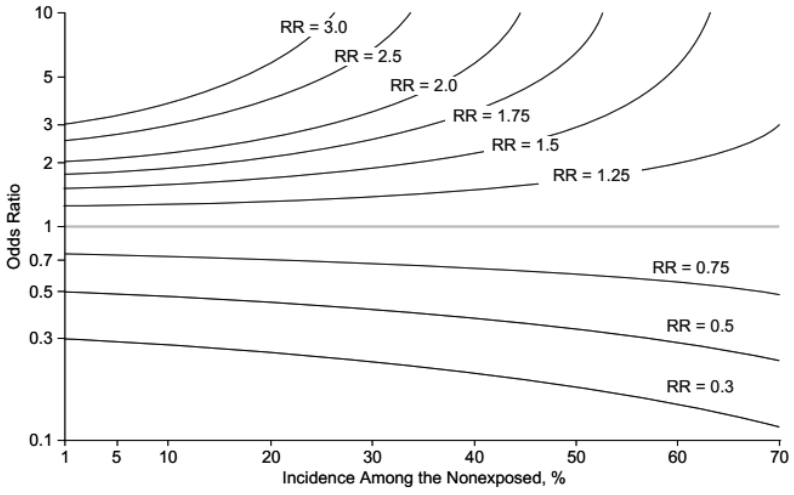


Figure I. 9 – The relationship between relative risk (RR) and odds ratio (OR) by incidence of the outcome.

Source: (Zhang & Yu 1998)

I.3.1.3. Absolute risk difference

The absolute risk difference (ARD) is obtained by subtracting the risk in one group from the risk in the other, i.e. is the difference in event rates between two groups. According to Table I. 5, ARD = |P1-P2|. The ARD can also be obtained through the multiplication of the RRR by the risk in control group (P2), i.e. ARD = RRR x P2. Depending on if there is a reduction or an increase in risk of events in the treatment group compared to the control group, the ARD is called absolute risk (AR) reduction or absolute risk increase (ARI), respectively.

Using the example provided in Table I. 6, the AR reduction between treatment and control groups would be estimated at 0.08 (=0.20-0.12) in patients with moderate hypertension and 0.006 (=0.015-0.090) in patients with mild hypertension. The AR reduction in the latter population is trivial (0.6%) compared to patients with moderate hypertension (8%).

The RR, RRR and OR reflect the effects of an intervention in proportional terms, but preclude conclusions about the size of effects on an absolute scale. The ARD gives an impression about whether an effect may be clinically meaningful or not. However, the ARD may be difficult to interpret and incorporate in clinical practice because it is a dimensionless abstract number that may be not easily and immediately perceived (Laupacis, Sackett & Roberts 1988), (McQuay & Moore 1997).

I.3.2. THE NUMBER NEEDED TO TREAT CONCEPT

The concept of number needed to treat (NNT) was introduced in the medical literature by Laupacis et al. in 1988. They aimed to propose a yardstick that the practicing clinician could use to measure and compare the benefits and risks of medical interventions (Laupacis, Sackett & Roberts 1988).

Such a yardstick would verify four properties: first, it would compare the consequences of doing nothing (i.e. the risk for an adverse event if no treatment is given) with the potential benefits of doing something (i.e. the reduction of risk provided by the intervention); second, it would express the harm associated to the treatment (e.g. adverse events and toxicity to the patient); third, it would identify patients at high risk for an event and responsive to therapy; fourth, it would consist of a measure that would allow to compare the consequences of different interventions, being useful for individual clinicians to support their decisions (Laupacis, Sackett & Roberts 1988).

Laupacis et al. suggested that the reciprocal of the AR reduction would be a highly useful measure for clinicians – i.e. the NNT, expressing "the number of patients with a given disorder that a physician must treat in order to protect one of them from the disorder's potential consequences" (Laupacis, Sackett & Roberts 1988). In other words, the NNT is the number of patients needed to be treated with one therapy versus another for one patient to encounter an additional outcome of interest within a defined period of time (McQuay & Moore 1997).

These researchers used data from a randomized placebo-controlled trial, in which patients were followed for a fixed amount of time to observe a binary response (event/no event), in order to point out some disadvantages of relative effect measures, while highlighting potential advantages of using absolute effect measures, namely the NNT, to express the consequences of clinical interventions (Laupacis, Sackett & Roberts 1988).

In the case of the hypothetical study presented in Table I. 6, in which the risk decreased from 0.20 to 0.12 with treatment versus control in patients with moderate hypertension (RR = 0.6, RRR =0.4, and ARR 0.08), the NNT would be approximately 13 (the NNT is usually expressed as positive whole number, all decimals being rounded up) (Straus et al. 2011), (Schünemann et al. 2011). This result means that a clinician would need to treat 13 patients with the experimental treatment to prevent stroke from occurring in one patient during a given period of time. It is important to recall that the RR and RRR was the same for patients with moderate and mild hypertension, but that the AR decreased from 0.20 to 0.12

in the first case and from 0.015 to 0.009 in the second case. The NNT would be estimated at approximately 167 for patients with mild hypertension receiving treatment as compared to those in the control group. Thus, a clinician would need to treat 167 patients with mild hypertension to prevent one stroke, but only 13 patients with moderate hypertension to obtain the same therapeutic result. The clinical recommendation is therefore probably different for these groups of patients (Cook & Sackett 1995).

The NNT may be advantageous over relative effect measures because it expresses clinical results in a manner that incorporates both the baseline risk without therapy and the risk reduction with therapy (Laupacis, Sackett & Roberts 1988). In addition, the NNT is more useful than AR reduction because it has the clinical immediacy of telling clinicians and patients how much effort is needed to achieve a particular therapeutic outcome (Laupacis, Sackett & Roberts 1988), (McQuay & Moore 1997).

Further, the NNT can be used to express results on both beneficial and harmful outcomes. Note that, when analysing the beneficial effects of a given therapy, a negative NNT means that the intervention has a harmful effect. In such situation, the NNT has been called number needed to harm (NNH) (Straus et al. 2011), (McQuay & Moore 1997). However, the NNT terminology to represent benefits and harms is not consensual. Altman suggested that it is more appropriate that the number of patients needed to be treated for one additional patient to benefit or be harmed should be denoted NNTB and NNTH, respectively (Altman 1998). In an ideal scenario, a particular treatment would have low values of NNTB ($\simeq 1$) and high values of NNTH ($\simeq \infty$), meaning that very few patients need to be treated to achieve clinical benefit, and a very high number of patients need to be treated for a harmful event to occur over a defined period of time.

I.3.2.1. Characteristics of the number needed to treat concept

There are some characteristics that are inherently associated with the concept of the NNT. The NNT refers to a specific comparator, a particular clinical outcome, and a given period of time. In order to be fully interpretable, these features should always be specified. In addition, as with other estimates of treatment effect, confidence intervals should be presented for the point-estimate NNT (Altman 1998). The NNT is therefore specific to a given comparison, rather than an isolated measure of effect specific to a single clinical intervention (McQuay & Moore 1997), (McAlister 2008).

There are three main factors that can influence the NNT, which are the baseline risk, the time horizon and, obviously, the outcome of interest.

I.3.2.1.1. Baseline risk

The relative effect of clinical interventions is usually similar across populations with varying baseline risks, particularly when interventions aim to modify risk factors and slow the progress of a disease (Schmid et al. 1998), (Sackett 2001), (Furukawa, Guyatt & Griffith 2002), (McAlister 2002). Assuming a constant RRR across a range of baseline risks for a given adverse outcome, the NNT varies inversely with baseline risk, as illustrated in Table I. 7 (Laupacis, Sackett & Roberts 1988). Thus, the NNT to prevent an adverse outcome seems, usually, less favourable in low-risk populations (McAlister 2008).

Table I. 7 – The effect of baseline risk and relative risk reduction on the number needed to treat.

Baseline risk*	Relative risk reduction by a new therapy (%)						
	50	40	30	25	20	15	10
	Number needed to treat						
0.9	2	3	4	4	6	7	11
0.6	3	4	6	7	8	11	17
0.3	7	8	11	13	17	22	33
0.2	10	13	17	20	25	33	50
0.1	20	25	33	40	50	67	100
0.05	40	50	67	80	100	133	200
0.01	200	250	333	400	500	667	1000
0.005	400	500	667	800	1000	1333	2000
0.001	2000	2500	3333	4000	5000	6667	10000

*Risk of an adverse event in control patients.
Source: (Laupacis, Sackett & Roberts 1988).

As illustrated in Table I. 6 for the hypothetical clinical trial, the NNT to prevent stroke is more favourable for patients with moderate hypertension, i.e. with higher baseline risk (NNT=13), than in patients with mild hypertension (NNT=167). Moreover, the NNT is likely to be influenced by secular changes of baseline risks, which tend to improve over time due to, for example, more timely diagnoses and more efficacious standard therapies. The NNT may be higher in recent trials compared to earlier ones because of patients having lower baseline risks for adverse outcomes in the present as compared to the past (McAlister 2008).

1.3.2.1.2. Time of follow-up

The length of time during which the patients are followed also influences the NNT, because the concept is inherently dependent on time. If an intervention produces a constant RRR, the NNT to prevent an adverse outcome tend to become more favourable (i.e. decrease) with increasing time of follow-up, because the events accumulate, and consequently the absolute event rates increase. As shown in Table I. 8, McAlister analysed the influence of time on NNT, by using data from a clinical trial with statins to prevent myocardial infarction in patients with hypertension (Sever et al. 2005), (McAlister 2008).

Table I. 8 – Influence of time horizon on the number needed to treat.

Time	Event rate in control group	Event rate in intervention group	RRR (95% CI)	AR Reduction (95% CI)	NNT (95% CI)
90 days	21/5121	7/5184	0.67 (0.23-0.86)	0.28 (0.07-0.48)	364 (210-1362)
12 months	61/5121	34/5184	0.45 (0.16-0.64)	0.54 (0.17-0.92)	186 (109-601)
3.3 years	154/5121	100/5184	0.36 (0.18-0.50)	1.08 (0.48-1.69)	93 (59-208)

CI, confidence interval.
Source: (McAlister 2008).

A formula has been proposed to extrapolate the NNT from one interval of time (*t*) to another (*s*) – NNT*t* x *t*/*s* = NNT*s*, where NNT/t and NNT/s are the numbers of persons needed to treat for time *t* and *s*, respectively (Laupacis, Sackett & Roberts 1988). As an example, if the results of a trial with one-year of follow up (NNT*t*=10) were extrapolated to a five-year horizon, the NNT would be estimated at 2 (NNT*s* =10 x 1/5). However, this type of extrapolations is usually not recommended because it can lead to biased estimates. The problem is that the formula assumes that benefits and harms, as well as RRR, remain constant over time, but this is usually not the case. For example, hydroxychloroquine starts reducing the risk of diabetes mellitus in patients with rheumatoid arthritis (baseline risk ≃0.09) after two years of treatment (RR, 0.76; RRR, 24%; NNT, 47), and continues to decrease with longer duration of >4 years (RR 0.69; RRR, 31%; NNT, 36) (Ozen et al. 2016). Using the converting formula, the NNT would be estimated at 24 (NNT=47 x 2/4) after 4 years of follow-up, i.e. overestimating the effect. Thus, this formula should be avoided in most cases. It is preferable to use, when available, survival curves to estimate event rates and apply hazard ratios to control event rates at times of interest (Altman & Andersen 1999), (Suissa et al. 2012).

I.3.2.1.3. Outcomes

A NNT is specific to a given outcome of interest. Since most therapies produce impact over several outcomes, usually more than one NNT needs to be calculated and incorporated for making clinical decisions. Clinicians should take into account the relative weigh of importance of each outcome and patient preferences when using NNT for supporting clinical decisions (McAlister 2008).

In general, the NNT for harmful outcomes (NNTH) should be higher than the NNT for beneficial outcomes (NNTB). This means that benefits are encountered more frequently than harms. Nevertheless, acceptable values of NNT (NNTB and NNTH) depend on the outcome of interest because different outcomes have different clinical importance. A single-digit NNTH may be acceptable if the outcome is an adverse event that is mild and transient, but a NNTH >1000 is probably necessary to accept the risk of a serious adverse event that may pose a significant health risk for the patient (Citrome & Ketter 2013). Further, the acceptance of a given NNTH depends not only on the nature of the harmful outcome, but also on the beneficial outcome that is achieved with the treatment, and the resulting NNTB, as well as the condition or disease being treated.

Noteworthy, the ratio (1 / NNTB) / (1 / NNTH) (or simply, NNTH/NNTB), called the likelihood of being helped or harmed (LHH) can be calculated to illustrate trade-offs between benefits and harms and to inform clinicians about how many patients might benefit from treatment for each one who is harmed. In case of LHH >1, the expected benefits outweigh possible harms (Citrome & Ketter 2013). Though, a high LHH is usually required when comparing a desired outcome with an adverse event that causes significant harm to the patient; and a low values of LHH may be acceptable if the adverse event is mild and transient (Citrome & Ketter 2013). Nevertheless, NNTB and NNTH values are often presented separately and less frequently as NNTH/NNTB ratios because investigators may be reluctant to weigh benefits and harms equally on the same scale given the uncertainties about their relative importance (Boyd et al. 2012).

I.3.3. INTERPRETATION OF CONFIDENCE INTERVALS FOR THE NUMBER NEEDED TO TREAT

The confidence interval for the NNT is usually calculated by taking the reciprocal of the values defining the confidence interval of the ARD (Cook & Sackett 1995). When there

is no treatment effect, the ARD is zero and the NNT is infinite. This situation may cause problems in the interpretation of the NNT (Altman 1998).

If the ARD is statistically significant at a 5% level, the 95% confidence interval will not include zero and the NNT will not include infinity. However, when the treatment effect is not statistically significant at the *p* threshold of <0.05, the 95% confidence interval for the ARD will include zero, and the 95% confidence interval for the NNT will include infinity (Altman 1998).

As illustrated by Altman, a treatment providing an ARR of 10% with a 95% confidence interval from –5% to 25% would result in a NNT of 10 and a 95% confidence interval from –20 to 4, which excludes the point-estimate (Altman 1998). This confidence interval apparently encompasses two disjoint regions, with values of NNT going from 4 to ∞ and NNT going from –20 to –∞ (Figure I. 10).

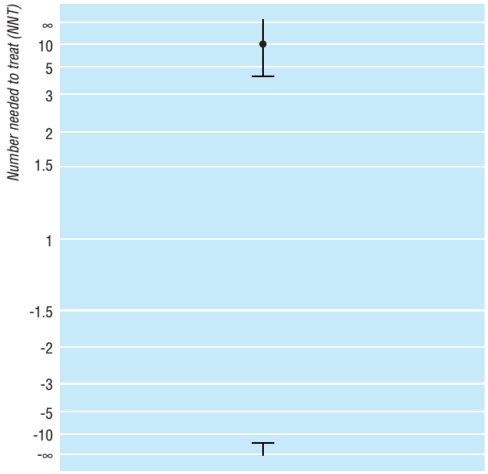


Figure I. 10 – Example: illustration of a 95% confidence interval for NNT=10.

Source: (Altman 1998).

Altman proposed an alternative way of representing and interpreting the confidence interval for the NNT. Of recall, this author proposed using NNTB and NNTH for representing benefits and harms, respectively (i.e. a negative value of NNTB is a positive value of NNTH) (Altman 1998).

Taking into account the AR reduction scale, that goes from –100% to 100% through zero, the NNT (=1/AR reduction) scale goes from NNTH = 1 to NNTB = 1 via infinity (∞) (Altman 1998). Using the example provided before, the 95% confidence interval for the NNTB 10 could be quoted as NNTH 20 to ∞ to NNTB 4 (Figure I. 11).

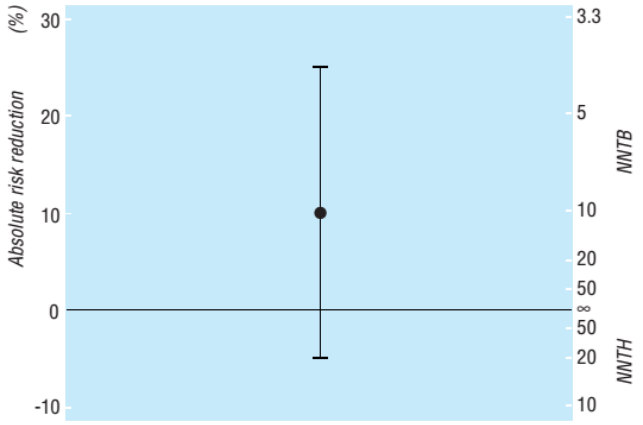


Figure I. 11 – Relation between absolute risk reduction, number needed to treat and their confidence intervals.

NNTB, number needed to treat to benefit; NNTH, number needed to treat to harm.
Source: (Altman 1998).

I.3.4. CALCULATION OF THE NUMBER NEEDED TO TREAT IN DIFFERENT STUDY DESIGNS

As previously discussed, the concept of NNT was introduced in the medical literature to express differences between treatments on binary outcomes in the context of RCTs with fixed times of follow-up (Laupacis, Sackett & Roberts 1988), (Cook & Sackett 1995). However, evidence on treatment effects may come from different sources of evidence, including various research designs and not only individual RCTs (e.g. meta-analysis, cohort and case-control studies). Further, the effects of clinical interventions may be expressed with non-binary variables (e.g. time to event outcomes – survival data). Other issues that warrant precaution when calculating and interpreting the NNT include for example studies with incomplete follow-up of patients, and events that occur repeatedly over time (e.g. exacerbations of asthma).

Clinicians, regulators and other stakeholders involved in the benefit-risk assessment of medicines should be aware that there are different approaches for calculating the NNT. In order to produce good estimates of the NNT (and other related or similar metrics), it is important to apply methods that are appropriate to the research question and the context of analysis, including the design of studies used as source of evidence, type of variables used to express outcomes of interest and other characteristics specific to each study.

This section describes the methods that are recommended for calculating the NNT in different scenarios, according to the evidence published in scientific literature.

I.3.4.1. Systematic reviews and meta-analyses

Systematic reviews and meta-analyses, particularly those including RCTs producing effect estimates with narrow confidence intervals, are frequently classified as the top level of evidence for supporting decisions in clinical research and practice (Berlin & Golub 2014).

The main goal of a systematic review is to collect and evaluate all relevant studies on a particular topic (Chalmers & Altman 1995). Systematic reviews are often conducted prior to a meta-analysis. A meta-analysis consists in a statistical analysis of a collection of analytic results, which purpose is to integrate the finding (Berlin, Soledad-Cepeda & Kim 2012).

For purposes of clinical research, meta-analyses produce effect size estimates that quantify a relationship between two variables or a difference between two groups (Borenstein et al. 2009). The effect size measures more commonly used in meta-analyses depend on the summary data reported in primary studies, and include OR, RR or risk difference (RD) for binary outcomes (i.e. having vs. not having an event); raw difference in means or standardized mean difference (SMD) for means and standard deviations; and hazard ratio (HR) for time to event outcomes, also called survival analysis (i.e. when the outcome of interest is assessed as the time elapsing before an event is experienced) (Borenstein et al. 2009), (Higgins & Green 2011). Of note, HR is interpreted similarly to RR; however, hazard is slightly different from risk as the first measures instant risk and may change continuously (Higgins & Green 2011).

As discussed earlier, the NNT calculated by taking the reciprocal of the ARD between two groups for a binary outcome. However, this calculating approach can be precluded in the context of meta-analyses because they summarize treatment effects in various ways (Deeks, Higgins & Altman 2011).

First, is important to recall that in a meta-analysis, the data for each study is summarised, and then those summaries are statistically combined and presented as a meta-analytical estimate. Treating data as it all come from a single trial (i.e. adding together raw totals of patients from each study) is not a valid approach for meta-analysis and should not be used to calculate NNT, because of Simpson's paradox (Cates 2002), (Altman & Deeks 2002).

In addition, although the NNT can be directly derived from meta-analyses presenting pooled RDs, this approach is usually not recommended and may result in biased estimates. The calculation of the NNT assumes that RDs are constant across trials. However, this is unlikely to be the case of most meta-analyses because of the inclusion of studies with various

baseline risks (i.e. different event rates in control groups), as well as different durations of follow-up. These issues influence the pooled ARD, its reciprocal, and consequently the NNT (Smeeth, Haines & Ebrahim 1999), (Cates 2002), (Altman & Deeks 2002), (Furukawa, Guyatt & Griffith 2002), (Marx & Bucher 2003), (Deeks, Higgins & Altman 2011).

Smeeth and colleagues argued that there is no single, true pooled ARD, as assumed in the fixed effects model, neither the variation in the RD between trials solely results of a sampling effect, as presumed in the random effects model for meta-analyses. In addition, pooled ARDs using number of patients as denominator assume identical duration of follow-up across trials, which is often not the case (Smeeth, Haines & Ebrahim 1999). Thus, when there is a high variance of RDs and baseline event rates across trials included in a meta-analysis, the NNT derived from a pooled ARD is not very informative and is possibly misleading (Marx & Bucher 2003).

In meta-analyses, the NNT should preferably be calculated using pooled estimates expressed as relative effects, rather than as absolute effects (Smeeth, Haines & Ebrahim 1999), (Cates 2002), (Marx & Bucher 2003), (Deeks, Higgins & Altman 2011). The available evidence suggest that, in general, OR and RR are more stable across different baseline risks compared to estimates of ARD (Schmid et al. 1998), (Engels et al. 2000), (Furukawa, Guyatt & Griffith 2002).

Under the assumption that the relative benefits and risks of therapy are the same regardless of the baseline risk, the NNT can be individualized for each patient in clinical practice using estimates of relative effects and the PEER (Smeeth, Haines & Ebrahim 1999), (McAlister et al. 2000), (Furukawa, Guyatt & Griffith 2002), (Cates 2002), (Marx & Bucher 2003), (Deeks, Higgins & Altman 2011), (Straus et al. 2011). Furukawa and colleagues found that point estimates of individualized NNT agree well, and are unlikely to cause divergent clinical decisions, across a range of values of PEER, when calculated from fixed effects OR, random effects OR and random effects RR (Furukawa, Guyatt & Griffith 2002).

Formulas to convert OR and RR to NNT are provided in Table I. 9. Confidence intervals for NNT can also be calculated by applying the same formulas to the upper and lower confidence limits for the summary statistic (i.e. RR or OR) (Altman 1998), (Cates 2002), (Deeks, Higgins & Altman 2011). However, this confidence interval does not incorporate uncertainty around the control event rate (Deeks, Higgins & Altman 2011).

Table I. 9 – Formulas to convert OR and RR to NNT.

	Formula
For RR <1	
	NNT = 1 / (1 – RR) x PEER
For RR >1	
	NNT = 1 / (RR – 1) x PEER
For OR <1	
	NNT = 1 – [PEER x (1 – OR)] / [(1 – PEER) x (PEER) x (1 – OR)]
For OR >1	
	NNT = 1 + [PEER x (OR – 1)] / [(1 – PEER) x (PEER) x (OR – 1)]

NNT, number needed to treat; OR, odds ratio; PPER, patient-expected event rate; RR, relative risk.
Source: (Straus et al. 2011)

Although meta-analyses are often used to pool overall estimates of risk for binary outcomes, they can also analyse outcomes that are measured on a continuous scale (e.g. intensity of pain, functional capacity). The most common approach is to generate a SMD (da Costa et al. 2012). However, this effect size measure is sometimes non-intuitive and difficult to interpret by clinicians (Thorlund et al. 2011). For that reason, results on continuous scales are often dichotomized using a responder analysis approach (i.e. patients are classified as responders and non-responders depending on the level of reduction in symptoms) (Farrar, Dworkin & Max 2006), (Henschke et al. 2014), (Falk et al. 2014). The dichotomized data can then be used to estimate differences between groups, applying OR, RR, RD or NNT (da Costa et al. 2012).

There are methods to convert SMDs or means to measures of dichotomized treatment response (Hasselblad & Hedges 1995), (Cox & Snell 1989), (Suissa 1991), (Kraemer & Kupfer 2006), (Furukawa & Leucht 2011). The authors of a study that analysed the performance of those methods concluded that four methods are suitable to convert summary treatment effects of continuous outcomes into OR and NNT (Hasselblad & Hedges 1995), (Cox & Snell 1989), (Suissa 1991), (Furukawa & Leucht 2011), (da Costa et al. 2012).

I.3.4.2. Randomized controlled trials

There is a number of issues that must be considered to calculate the NNT in the context of RCTs. The classical approach to calculate the NNT, i.e. the reciprocal of the ARD, works well if the RCT assesses binary outcomes and all patients complete a pre-defined fixed time of follow-up. However, there are for example RCTs that assess the effect of interventions on time to event outcomes, such as rates of survival, and therefore it is necessary to consider the influence of varying follow-up times in the estimation of the NNT

(Altman & Andersen 1999), (Suissa et al. 2012), (Bender et al. 2013). The approaches to be used in several scenarios are further discussed.

I.3.4.2.1. Time to event outcomes and varying times of follow-up

The calculation of the NNT is more challenging and more prone to bias in RCTs where the outcome is time to event (Mayne, Whalen & Vu 2006), (Suissa et al. 2012), (Bender et al. 2013), (Suissa 2015). A review of trials assessing such outcomes found that only 50% applied appropriate calculating methods (Hildebrandt, Vervölgyi & Bender 2009).

The calculation of the classical person-based NNT is founded on the cumulative incidence of the outcome per number of patients followed over a defined period of time. Thus, in studies with varying times of follow-up, the calculation of the proportions of patients with the outcome of interest must be adjusted to this time variations (Suissa et al. 2012), (Suissa 2015). Two calculating methods have been proposed with the aim to adjust NNT estimates in studies with varying follow-up.

I.3.4.2.1.1. Survival probabilities: Kaplan-Meier approach

The first method is based on survival probabilities obtained by means of the Kaplan-Meier survival curves or the Cox regression model (Altman & Andersen 1999). Instead of using simple proportions, a Kaplan-Meier approach is recommended to estimate correct proportions because it accounts for varying times of follow-up and provides a curve of cumulative incidences over time (Collet 1994), (Suissa et al. 2012), (Suissa 2015). This method allows calculating the NNT for a specific point of time of follow-up, representing the number of patients needed to treat so that one more patient is free of the event in the treatment group compared to the control group at that point of time (Altman & Andersen 1999).

Depending on the information that is available, the NNT and its confidence intervals can be calculated based on two approaches (Altman & Andersen 1999). First, if only survival probabilities are available, i.e. Kaplan-Meier curves have been generated, the NNT can be obtained as follows: $NNT = 1 / (S_a - S_c)$, where S_a and S_c are the estimated survival probabilities for active treatment and control group, respectively, at a given point of time (Altman & Andersen 1999). Second, if the only information available is about the survival probability in the control group and the estimate of HR, then the $NNT = 1 / [S_c(t)]^h - S_c(t)]$,

in which $S_c(t)$ is the survival probability for control group at time t, and h is the HR comparing the two groups (Altman & Andersen 1999). Formulas to calculate confidence intervals are available elsewhere (Altman & Andersen 1999).

Suissa used a hypothetical scenario to illustrate that the lack of adjustment to varying follow-up times, i.e. using simple proportions, may result in distorted values of the NNT (Suissa 2015).

In a hypothetical RCT, 3000 patients with iron overload syndrome would be assigned (1:1:1) to one of three groups (placebo, feclad or fedom) and followed for one year or until liver failure (Suissa 2015). However, 60% of patients were censored before one year of follow-up (mean follow-up was 7 months). The cumulative incidence curves (the reverse of Kaplan-Meier curves) of liver failure are presented in Figure I. 12.

Table I. 10 shows significant differences between the values of NNT obtained from simple proportions and those obtained using a proper Kaplan-Meier approach (Suissa 2015).

Simple proportions should not be used to estimate NNT in RCTs unless all patients are followed for the full study duration – this is the only situation in which simple proportions produce the same cumulative incidences as those obtained by a Kaplan-Meier approach. Nevertheless, the two approaches produce similar estimates of cumulative incidences in RCTs with short and mostly complete times of follow-up (Suissa 2015).

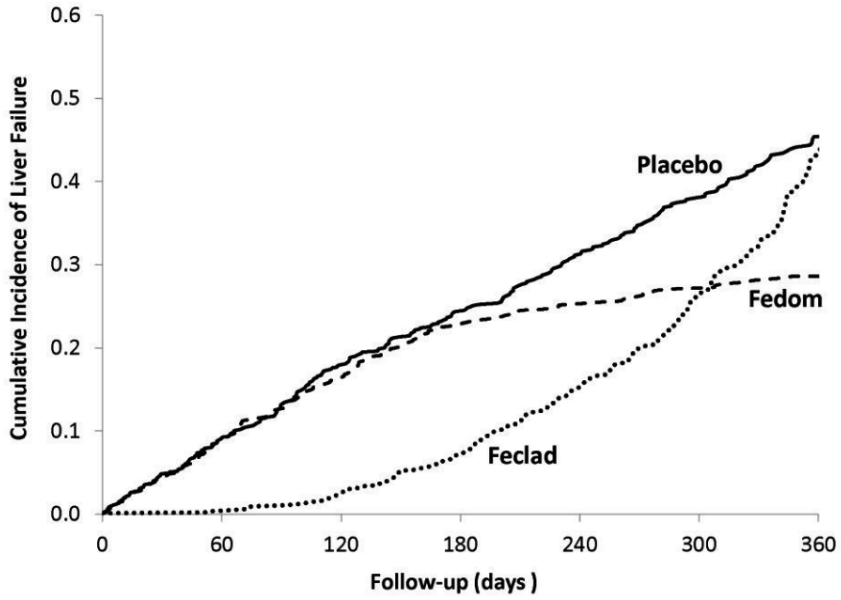


Figure I. 12 – Cumulative incidence of liver failure in a hypothetical RCT.

Source: (Suissa 2015).

Table I. 10 – Comparison between NNT calculated from simple proportion with that from Kaplan-Meier approach in a hypothetical RCT.

Group	Patients, N	Patients with liver failure, N	Kaplan-Meier approach		Simple proportion	
			1-year cumulative incidence	NNT	1-year cumulative incidence	NNT
Placebo	1000	324	0.454		0.324	
Fedom	1000	230	0.286	6	0.230	11
Feclad	1000	238	0.441	77	0.238	12

Source: (Suissa 2015).

I.3.4.2.1.2. Incidence rates per person-time

The other method is based on the reciprocal of the difference of annualized incidence rates, rather than on the reciprocal of the difference of absolute risks (Lubsen, Hoes & Grobbee 2000), (Mayne, Whalen & Vu 2006).

The authors argued that the calculation of the NNT would be more appropriate using units of person-time for chronic conditions that require continuous treatment. One of the arguments was that the classical NNT (i.e. the reciprocal of ARD) decrease as function of time if the relative risk between groups remains constant while events accrue over time. They suggested that such situation may lead to the misleading conclusion that the effectiveness of therapy improves over time. However, they note that specifying the point of time for which the classical NNT is calculated helps clarifying its interpretation (Mayne, Whalen & Vu 2006).

Incidence rates are calculated by dividing the number of patients with the outcome of interest by the total number of person-time of follow-up. The person-time based NNT is given by $1 / (IR_0 - IR_1)$, where IR is incidence rate, 0 represents control group and 1 the treatment under evaluation (Mayne, Whalen & Vu 2006), (Suissa et al. 2012), (Bender et al. 2013).

This calculating method estimates the number of person-time (e.g. patient-years), not the absolute number of persons, needed to observe one less event in the treatment group than in the control group (Mayne, Whalen & Vu 2006), (Suissa et al. 2012), (Bender et al. 2013). Thus, this method will result in an estimate that is different from the classical person-based NNT and may be difficult to interpret (Bender et al. 2014). Of note, 100 patient-years do not necessarily mean 100 individual patients treated over one year (or 50 patients treated for two years). Examples of incorrect interpretations of person-time based NNT is provided elsewhere (Suissa et al. 2012).

Analysing again the example provided by Suissa, the use of simple incidence rates per patient-year also results in different values of NNT compared to the proper Kaplan-Meier approach, namely for feclad versus placebo (Table I. 11) (Suissa 2015).

Table I. 11 – Comparison between NNT calculated from simple incidence rates per person-time with that from Kaplan-Meier approach in a hypothetical RCT.

Group	Patients, N	Patients with liver failure, N	Kaplan-Meier approach		Incidence rate per patient-year	
			1-year cumulative incidence	NNT	Incidence rate per patient-year	NNT
Placebo	1000	324	0.454		0.589	
Fedom	1000	230	0.286	6	0.399	6
Feclad	1000	238	0.441	77	0.387	5

Source: (Suissa 2015).

Noteworthy, inverting differences of incidence rates as a measure of effect to express amount of person-time is only valid in the case of a constant hazard difference, i.e. the distribution of the survival times follow the exponential distribution or the linear hazard rate distribution (Lin, Wu & Balakrishnan 2003), (Mayne, Whalen & Vu 2006), (Stang, Poole & Bender 2010), (Suissa et al. 2012), (Bender et al. 2013), (Bender et al. 2014).

The authors of a more recent study that compared the two methods, i.e. inverting ARDs obtained by survival time approaches and inverting incidence rates differences, concluded that the second method is more prone to bias and low coverage properties across a wide range of data situations. They recommended the use of ARDs to estimate NNT in RCTs with time to event outcomes (Bender et al. 2013).

I.3.4.3. Observational studies

The concept of NNT has been applied in epidemiological research, including case-control and cohort studies (Bjerre & LeLorier 2000), (Heller et al. 2002), (Bender & Blettner 2002). However, other terminologies have been proposed considering that the factor under evaluation may be exposition rather than treatment. As such, the concept of NNT can been designated as "number needed to be exposed" (NNE) (Bender & Blettner 2002), or "exposure impact number" (EIN) (Heller et al. 2002) in the context of epidemiological research. Irrespectively of the terminology, the principle is the same, i.e. the reciprocal of RD between groups (Bender et al. 2007).

Bender and colleagues applied the NNE to describe the average number of unexposed persons needed to be exposed to observe one extra case; and the EIN to describe the average number of exposed persons amongst whom one excess case is due to the exposure (Bender et al. 2007). In addition, they proposed using NNEH (NNEB) to describe the number needed to be exposed for one person to be harmed (benefit) (Bender & Blettner 2002), (Bender et al. 2007).

Due to the lack of randomization in observational studies, the baseline characteristics of exposed individuals (e.g. treatment or a risk factor) may differ systematically and significantly from those of unexposed ones (Grimes & Schulz 2002a), (Trojano et al. 2017).

In order to avoid biased results, adjusting for confounding covariates (e.g. regression-based methods or propensity score methods) is required before producing estimates about the effect of exposition on outcomes (Bender 2009), (Haukoos & Lewis 2015), (Trojano et al. 2017). Although this adjustment is routinely carried out to estimate relative effects (e.g. OR, HR) between groups, NNT measures are often obtained from crude RDs without adjustment for confounding factors (Bender & Blettner 2002). The calculation and interpretation of unadjusted estimates of NNT measures may lead to misleading conclusions.

The calculation of NNT measures in observational studies is further discussed with respect to research designs, adjustment approaches and outcomes.

I.3.4.3.1. Case-control studies

Case-control studies aim to investigate if there are differences in previous exposures (e.g. treatments or risk factors) between cases with a given outcome and controls without that outcome (Rosenberg, Coogan & Palmer 2012), (Strom 2013). Such studies are commonly conducted to investigate risks of rare adverse events or diseases of long term latency (Grimes & Schulz 2002b).

This research design precludes the estimation of incidence rates, unless the study is nested within a cohort (Bjerre & LeLorier 2000), (Grimes & Schulz 2002b). Estimates of risk are usually expressed by means of OR, which compares the proportion of exposed subjects among cases and controls (Grimes & Schulz 2002b).

Bjerre & LeLorier proposed using OR (and limits of its confidence interval) and unexposed event rate (UER) (i.e. with the same meaning as CER or PEER) to calculate NNT (with confidence interval) (Bjerre & LeLorier 2000). The UER can be estimated from

external sources, such as controls in RCTs or unexposed subjects in cohort studies (Bjerre & LeLorier 2000).

This approach allows to calculate adjusted NNTs by using ORs that are adjusted for confounding factors (Bender & Blettner 2002). Adjusted ORs are often obtained by means of logistic regression. The formula used to convert adjusted ORs into adjusted NNTs is provided in Table I. 9 (Bender & Blettner 2002).

I.3.4.3.2. Cohort studies

In cohort studies, two or more groups of patients are followed (retrospectively and/or prospectively) over time until the occurrence of a given outcome of interest. Usually these studies are used to compare a group of patients exposed to a risk factor (for example, a medicine) with an unexposed group of patients, or with a group of patients exposed to a different risk factor (Strom 2012), (Strom 2013). The groups are tested for differences between them in frequencies of outcomes of interest, and associations are possibly suggested (Grimes & Schulz 2002b).

Adjusted NNTs can be calculated in cohort studies by using adjusted ORs estimated, for example, by multiple logistic regression (Bender & Blettner 2002). Although the OR is constant over the distribution of considered confounders, the event rates and their differences vary with confounder values. Thus, NNT also varies depending on these values. This should be considered when adjusted NNTs are estimated using adjusted ORs and UERs (Bender & Blettner 2002). In cohort studies, the mean risk of unexposed persons (UER) can be estimated within the logistic regression framework for the corresponding confounder profile and then used to calculate an adjusted NNT. Alternatively, adjusted NNT can be calculated for some fixed confounder profiles (Bender & Blettner 2002).

Another approach was later proposed by two independent authors to calculate adjusted NNTs in cohort studies (Bender et al. 2007), (Austin 2010). This approach considers the distribution of confounders by using average RD estimated from logistic regression analysis. There are minor variations between the approaches suggested by the two authors. While Bender et al. suggested averaging the predicted probabilities over either the treated subjects or the untreated subjects (Bender et al. 2007), Austin suggested averaging the predicted probabilities across the entire sample (Austin 2010). The adjusted NNT is then calculated by taking the reciprocal of the average RD (Bender et al. 2007).

The average RD approach was considered to be better than the OR approach in terms of bias and coverage probability, particularly when the distribution of the confounders is wide (Bender et al. 2007). Nevertheless, the OR approach still leads to reasonable results in case of continuous confounders with narrow variability. In case of a wide distribution of the confounders, the OR approach may lead to a downward bias of NNT, i.e. an overestimation of the effect (Bender et al. 2007).

In cohort studies where the outcome is time to event, NNT can be estimated as the reciprocal of the average RD for a given duration of follow-up obtained from an adjusted survival model, such as the Cox proportional hazards regression model (Austin 2010), (Laubender & Bender 2010). This approach is similar to that one described above for estimating average RDs within the logistic regression model (Austin & Laupacis 2011). This approach allows to obtain NNT measures for survival data adjusted for confounders (Laubender & Bender 2010).

In addition to regression-based approaches, propensity score methods can be used to produce effect estimates with adjustment for confounders. The confounding between treatment status and baseline covariates is eliminated by matching or stratifying on propensity score, or weighting by the inverse probability of treatment (Austin & Laupacis 2011). Thus, the design of an observational study can be separated from the analysis of an observational study (Rubin 2007). Usually, outcomes can be directly compared between treated and untreated subjects without further adjustments for baseline covariates, i.e. as in RCTs (Austin & Laupacis 2011).

The NNT can be calculated as the reciprocal of the RD, which is estimated directly by comparing the estimated probability of the binary outcome between treated and untreated subjects in the matched sample in propensity-score matching (Austin 2011), (Austin & Laupacis 2011).

For time-to-event outcomes in cohort studies using propensity score methods, the NNT can be calculated as the reciprocal of the ARD estimated from Kaplan-Meier survival curves in treated and untreated subjects within a given duration of follow-up (Austin & Laupacis 2011). Different approaches are used to compare Kaplan-Meier survival curves between treated and untreated subjects depending on the propensity score method that is used, i.e. matching on propensity score (Klein & Moeschberger 1997), stratifying on propensity score (Austin & Laupacis 2011), or weighting by the inverse probability of treatment (Xie & Liu 2005).

I.4. NUMBER NEEDED TO TREAT IN CLINICAL PRACTICE

I.4.1. EVIDENCE-BASED MEDICINE AND CLINICAL DECISION

Clinical decisions about the care of the individual patient should be made upon the use of current best evidence (Sackett et al. 1996). This is the principle of evidence-based medicine (EBM), which term was coined in 1992 by a group led by Gordon Guyatt (Evidence-Based Medicine Working Group 1992). The practice of EBM is about the integration of the best research evidence, the clinical expertise of the clinician, and the patient's unique values and circumstances (Haynes & Haines 1998). There are five key steps in EBM, namely the following: 1) converting the need for information (about prevention, diagnosis, treatment, etc.) into an answerable clinical question; 2) searching for the best evidence that provides answer to that question; 3) appraising that evidence for its validity (closeness to the truth), impact (size of effect), and applicability (usefulness in clinical practice); 4) integrating the critical appraisal with clinical expertise and with patient preferences, and applying it to practice; 5) self-evaluating the effectiveness and efficiency in executing the previous steps and seeking ways to keep improving this process and further decisions (Straus et al. 2011).

The practice of EBM is not a "one-size fits all" approach (Straus et al. 2011). It should rather imply a shared decision making process that involves, at least, the clinician and the patient (Barry & Edgman-Levitan 2012), (Stiggelbout et al. 2012). In this process, the clinician act on the appraised evidence and must be able to inform the patient about the benefits and harms of treatment options, as well as their relative effectiveness, and eventually their costs. The patient expresses individual preferences and values (Barry & Edgman-Levitan 2012), (Oshima-Lee & Emanuel 2013). Each intervenient possesses therefore a better understanding of the pertinent factors and shares responsibility in the decision to be adopted (Charles, Gafni & Whelan 1997).

The authors of a systematic review concluded that clinicians rarely have accurate expectations of the benefits and harms of medical interventions. Although inaccuracies are seen in both directions, clinicians tend often to overestimate benefits and underestimate harms (Hoffmann & Del Mar 2017). This finding is in favour of the existence of therapeutic illusion in some cases, that is an unjustified enthusiasm with regards to a given treatment (Casarett 2016). Moreover, clinicians may have a tendency to search in a selective way for evidence that supports interventions they already use and consider to be effective, possibly

resulting in "confirmation bias" (Casarett 2016). For example, the results of a survey of urologists and radiation oncologists about the treatment of prostate cancer indicated that specialists overwhelming recommend the intervention that they themselves deliver (Fowler et al. 2000). In opposition, clinicians are less likely to recommend an intervention for which they have high expectations of harm (Gross et al. 2003), (Murthy, Kauldher & Targownik 2006). If clinicians have inaccurate perceptions about the benefits and harms of medical interventions, informed decision making, as well as optimal patient care may be compromised (Hoffmann & Del Mar 2017).

The clinician must understand the magnitude of benefits and harms that are potentially delivered to the patient by different therapeutic interventions. The balance between favourable and unfavourable effects of therapeutic alternatives is a necessary condition for making informed clinical decisions, such as determining clinical recommendations, or developing treatment guidelines (Laine, Taichman & Mulrow 2011).

The magnitude of treatment impact on outcomes of interest may be expressed using either relative or absolute measures of effect. The judgement exercised by clinicians to support clinical decisions is influenced by the format of presentation of the treatment effects, i.e. the statistical framing (McGettigan et al. 1999), (Nexoe et al. 2002), (Akl et al. 2011). The analysis of relative effects in isolation may lead to misleading conclusions. Clinicians may overestimate the magnitude of treatment effects if the results are expressed only in relative terms (Forrow, Taylor & Arnold 1992), (Naylor, Chen & Strauss 1992), (McGettigan et al. 1999), (Nexoe et al. 2002). For example, clinicians are more likely to prescribe a medicine that provides a 50% relative risk reduction of death than a medicine that reduces the absolute risk of death from 2% to 1%, or that need to be used by 100 patients to prevent one death. Nevertheless, these three statistical representations (RRR, AR reduction, and NNT) express the same effect (Alonso-Coello et al. 2016). The exclusive use of absolute effects has also downsides, mainly because, unlike relative effect measures, they are not stable across different baseline risks (Schmid et al. 1998), (Furukawa, Guyatt & Griffith 2002), (Deeks, Higgins & Altman 2011). Therefore, using a single estimate of absolute risk reduction to express treatment impact may underestimate the effect in high-risk patients, or overestimate the effect in low-risk patients. For this reason, in meta-analysis it is recommended that pooled findings are expressed by means of relative effect measures. Though, absolute effects can be obtained through the application of pooled relative effects to a range of baseline risks in the population of interest (Deeks, Higgins & Altman 2011).

Presenting reductions or increases in absolute risks provides often more transparent information than reporting reductions or increases in relative risks (Gigerenzer, Wegwarth & Feufel 2010). For example, a 24% reduction of breast cancer mortality was reported with mammography screening (Larsson et al. 1996). This figure was incorrectly interpreted by a considerable proportion of clinicians. When 150 gynaecologists were questioned about the meaning of a 25% reduction in the risk of death for breast cancer, 31% of them answered that 25 or 250 fewer women would die for every 1000 who were screened. However, the figure actually corresponded to a reduction from five to four deaths in every 1000 women, i.e. one less woman would die for every 1000 (0.1%) going under screening (Gigerenzer et al. 2007).

A study, in which clinicians were randomly assigned to four statistical framing formats about the effects of a new versus an old medicine in a hypothetical controlled clinical trial, found that the proportions of clinicians judging the new medicine as more effective differed depending on the risk presentation format (absolute survival 51.8%, absolute mortality 68.3%, relative mortality reduction 93.8%, and all three presented 69.8%). Compared to the presentation of all formats together, the greater perceived effectiveness was noted with relative mortality reduction (OR 4.40, $p<0.001$). The least biased interpretation is given by absolute risk framing (Perneger & Agoritsas 2011). The authors of another study found that the decisions made by cardiologists varied by the presentation format of benefits, with a higher proportion of clinicians recommending the treatment when the results were presented as RRR (62.2%), compared to AR reduction (40.4%) or the NNT (44.4%) ($p<0.001$ for both comparisons). Interestingly, these cardiologists interpreted the statistical evidence in the same manner regardless of data had been presented as AR reduction or by means of the correspondent NNT ($p=0.073$). The authors concluded that these professionals tended to misinterpret clinical data presented by means of relative effect measures (Borracci, Piñeiro & Arribalzaga 2015).

The presentation of both relative and absolute effects has been encouraged and acknowledged as necessary to improve the interpretability of treatment effects (Gigerenzer, Wegwarth & Feufel 2010), (Perneger & Agoritsas 2011), (Froud et al. 2012), (Busse & Guyatt 2015), (Alonso-Coello et al. 2016). It has also been argued that the absolute difference is of utmost interest and should determine clinical decisions (Busse & Guyatt 2015). Moreover, reporting the baseline risk, as well as the risk under treatment for the outcome of interest is important to support informed medical decisions (Stovitz & Shrier 2013).

However, absolute effect estimates are usually poorly reported in individual studies and also in systematic reviews (Schwartz et al. 2006), (King, Harper & Young 2012), (Beller et al. 2011). In addition, one third of systematic reviews presents mismatched framing, i.e. using relative effect measures to express benefits and absolute effect measures to express harms of interventions (Sedrakyan & Shih 2007). This may give the impression of large benefits and small risks with the treatment. For example, while the benefit of treatment given by a reduction of the probability of disease from 10 to 5 in 1000 patients could be expressed as a 50% reduction, the possibility of increasing the risk of disease from five to ten could be interpreted as an increase of five in 1000, i.e. 0.5% (Gigerenzer, Wegwarth & Feufel 2010). Most systematic-reviews (63.9%) still do not report absolute effects. In addition, those that do present such estimates, often report them inadequately (Alonso-Coello et al. 2016).

There is a need to improve the reporting of treatment effects in medical literature, namely by the presentation of absolute effects, such as the ARD or the NNT. The NNT has been acknowledged as a useful metric to support clinical assessments, and also to be used as guideline for decision making, for example in rheumatology clinical practice, and in chronic pain management (Osiri et al. 2003), (Moore et al. 2008), (Moore et al. 2010), (Katz, Paillard & Van Inwegen 2015). Moreover, the NNT has been used to estimate the effectiveness of implementing guidelines in clinical practice (Egan et al. 2016). The use of this metric in the assessment of benefits and risks of clinical interventions deserves further evaluation.

I.4.2. APPLICATION OF THE NUMBER NEEDED TO TREAT IN CLINICAL PRACTICE

I.4.2.1. Determination of the benefit-risk ratio for the individual patient

The NNT reported in published evidence (e.g. clinical trial or meta-analysis) is not always directly applicable to an individual patient, for example because the baseline risk of the patient is different from the published one or the RRR varies across subgroups of patients. Clinicians should consider patient's unique characteristics, which may influence benefits and risks of therapy, as well as patient's values when integrating research evidence in supporting clinical decision marking (McAlister et al. 2000), (Straus et al. 2011).

The NNT can be used to estimate risks of outcomes of interest (both benefits and harms), which are specifically adjusted to the characteristics of patients treated in real clinical

practice. There are two approaches that can be used to accomplish this: generation of patient-specific baseline risks (also called the patient's estimated event rate [PEER]); or clinical judgement. In both approaches, the relative benefits and risks of therapy are assumed to be the same whether the patients have high or low PEERs (McAlister et al. 2000), (Straus et al. 2011).

The first approach is a more exhaustive process that begins with the estimation of the PEER for the individual patient using data from various sources. Then, the PEER is applied to the overall RRR or relative risk increase (RRI) for calculating NNTB and NNTH, respectively, specifically adjusted to the individual patient (NNTB = 1 / [PEER x RRR]; NNTH = 1 / [PEER x RRI]).

According to the second approach, the clinician use the NNT (NNTB or NNTH) reported in a clinical study to generate the patient's risk of the outcome event (i.e. if the patient received control intervention), which is relative to that of the average control patient, and expressed as a decimal fraction (labelled f_t). The patient-specific NNT is then calculated by dividing the average NNT by f_t. For example, if a patient is judged to have twice the risk of the outcome as compared to control patients from a study, f_t = 2. Thus, in a hypothetical study reporting a NNTB of 20 for preventing a stroke over five years with therapy versus control, the patient-specific NNTB would be 10 (NNTB/f_t = 20/2). The same principle is followed with adverse events for estimating patient-specific NNTHs.

I.4.2.2. Incorporation of patient values and preferences

The NNTB and NNTH are useful for clinicians, but can be less informative for patients because they are more interested in their individual risks. Using a LHH adjusted to patient specific characteristics, values and preferences may be useful in the context of clinical decision making. During the discussion between clinician and patient, the latter can point out his preferences and values about receiving a therapy, namely the perception on the severity of potential adverse events and the severity of events that the therapy intends to avoid (McAlister et al. 2000), (Straus et al. 2011).

Using the hypothetical clinical study presented before, the NNTB to avoid a disabling stroke over 5 years is 20 for the average patient receiving treatment; however, the treatment also results in an increased risk of major bleeding, with a NNTH estimated at 60. A first approximation of the LHH would be calculated as LHH = (1/NNTB) / (1/NNTH) =

(1/20) / (1/60) = 3 to 1 in favour of treatment. The patient can be told that the treatment is 3 times as likely to help him as harm him.

However, the preliminary approach would not take into account neither the individual characteristics of the patient nor his preferences or values. As discussed above, the patient have a higher risk of stroke, with f_t estimated at 2. Further, let's say that the patient has a risk factor that increases 3-times the risk of major bleeding from treatment (labelled f_h for harm). The adjusted LHH would be calculated as LHH = [(1/NNTB) x f_t] / [(1/NNTH) x f_h] = [(1/20) x 2] / [(1/60) x 3] = 2 to 1 in favour of treatment.

This second LHH still neglects the patient values and preferences. For example, the patient may consider that having major bleeding is 10 times worse than having a disabling stroke. A severity factor (s) can be used to adjust LHH as follows: LHH = [(1/NNTB) x f_t] / [(1/NNTH) x f_h x s] = [(1/20) x 2] / [(1/60) x 3 x 10] = 5 in favour of not receiving treatment.

I.5. OBJECTIVES OF THIS THESIS

The number needed to treat (NNT) can be useful as metric to quantitatively assess benefits (NNTB) and harms (NNTH) of medicines during their entire life-cycle, and therefore have the potential to help increasing the objectivity, transparency and reproducibility of benefit-risk assessments. The assessment of risks of medicines, that is the safety profile, is particularly challenging, namely in post-marketing.

The primary objective of this project is to identify the potential role of the NNT as a metric for benefit-risk assessment of marketed medicines.

The specific objectives outlined for this project were the following:

1) To investigate the usefulness of metric indices for post-marketing safety evaluations, by estimating NNTH values for cardiovascular adverse outcomes for rosiglitazone (withdrawn from the EU market due to safety reasons, but still marketed in the USA) and pioglitazone (the other thiazolidinedione).

2) To explore the usefulness of NNTH in post-marketing benefit-risk assessments, by studying the agreement between NNTH values and withdrawals of medicines from the market due to safety reasons, and therefore to assess whether the results are in line with regulatory authorities' decisions.

3) To test NNTB, NNTH and LHH as metrics to assess benefits, risks and benefit-risk ratios of medicines in a therapeutic area that is associated with challenging clinical decisions with respect to the selection of adequate treatments, given the recent growth of the therapeutic arsenal.

4) To evaluate whether the methods applied by researchers to calculate the NNT in clinical literature are in line with basic methodological recommendations.

To fulfil point 1), the study entitled "Number needed to harm in the post-marketing safety evaluation: results for rosiglitazone and pioglitazone" was conducted (Mendes, Alves & Batel-Marques 2015); to fulfil point 2), the study entitled "Testing the usefulness of the

number needed to treat to be harmed (NNTH) in benefit-risk evaluations: case study with medicines withdrawn from the European market due to safety reasons" was conducted (Mendes, Alves & Batel-Marques 2016a); to fulfil point 3), the study entitled "Benefit-Risk of Therapies for Relapsing-Remitting Multiple Sclerosis: Testing the Number Needed to Treat to Benefit (NNTB), Number Needed to Treat to Harm (NNTH) and the Likelihood to be Helped or Harmed (LHH): A Systematic Review and Meta-Analysis" was conducted (Mendes, Alves & Batel-Marques 2016b); to fulfil point 4), the study entitled "Number needed to treat (NNT) in clinical literature: an appraisal" was conducted (Mendes, Alves & Batel-Marques 2017).

I.6. REFERENCES

AKL, E. ET AL., 2011. FRAMING OF HEALTH INFORMATION MESSAGES. COCHRANE DATABASE SYST REV, VOLUME 12, P. CD006777.

AKOBENG, A., 2016. UNDERSTANDING TYPE I AND TYPE II ERRORS, STATISTICAL POWER AND SAMPLE SIZE. ACTA PAEDIATR, 105(6), PP. 605-609.

ALONSO-COELLO, P. ET AL., 2016. SYSTEMATIC REVIEWS EXPERIENCE MAJOR LIMITATIONS IN REPORTING ABSOLUTE EFFECTS. J CLIN EPIDEMIOL, VOLUME 72, PP. 16-26.

ALTMAN, D., 1998. CONFIDENCE INTERVALS FOR THE NUMBER NEEDED TO TREAT. BMJ, 317(7168), PP. 1309-1312.

ALTMAN, D. & ANDERSEN, P., 1999. CALCULATING THE NUMBER NEEDED TO TREAT FOR TRIALS WHERE THE OUTCOME IS TIME TO AN EVENT. BMJ, 319(7223), PP. 1492-1495.

ALTMAN, D. & DEEKS, J., 2002. META-ANALYSIS, SIMPSON'S PARADOX, AND THE NUMBER NEEDED TO TREAT. BMC MED RES METHODOL, 2(3).

ALVES, C., BATEL-MARQUES, F. & MACEDO, A., 2014. DRUG-SAFETY ALERTS ISSUED BY REGULATORY AUTHORITIES: USEFULNESS OF META-ANALYSIS IN PREDICTING RISKS EARLIER. EUR J CLIN PHARMACOL, 70(6), PP. 745-756.

ALVES, C., MACEDO, A. & BATEL-MARQUES, F., 2013. SOURCES OF INFORMATION USED BY REGULATORY AGENCIES ON THE GENERATION OF DRUG SAFETY ALERTS. EUR J CLIN PHARMACOL, 69(12), PP. 2083-2094.

ARNARDOTTIR, A. ET AL., 2011. ADDITIONAL SAFETY RISK TO EXCEPTIONALLY APPROVED DRUGS IN EUROPE? BR J CLIN PHARMACOL, 72(3), PP. 490-499.

AUSTIN, P., 2010. ABSOLUTE RISK REDUCTIONS, RELATIVE RISKS, RELATIVE RISK REDUCTIONS, AND NUMBERS NEEDED TO TREAT CAN BE OBTAINED FROM A LOGISTIC REGRESSION MODEL. J CLIN EPIDEMIOL, 63(1), PP. 2-6.

AUSTIN, P., 2011. OPTIMAL CALIPER WIDTHS FOR PROPENSITY-SCORE MATCHING WHEN ESTIMATING DIFFERENCES IN MEANS AND DIFFERENCES IN PROPORTIONS IN OBSERVATIONAL STUDIES. PHARM STAT, 10(2), PP. 150-161.

AUSTIN, P. & LAUPACIS, A., 2011. A TUTORIAL ON METHODS TO ESTIMATING CLINICALLY AND POLICY-MEANINGFUL MEASURES OF TREATMENT EFFECTS IN PROSPECTIVE OBSERVATIONAL STUDIES: A REVIEW. INT J BIOSTAT, 7(1), P. 6.

BANZI, R., GERARDI, C., BERTELE', V. & GARATTINI, S., 2015. APPROVALS OF DRUGS WITH UNCERTAIN BENEFIT-RISK PROFILES IN EUROPE. EUR J INTERN MED, 26(8), PP. 572-584.

BARRY, M. & EDGMAN-LEVITAN, S., 2012. SHARED DECISION MAKING — THE PINNACLE OF PATIENT-CENTERED CARE. N ENGL J MED, 366(9), PP. 780-781.

BELLER, E., GLASZIOU, P., HOPEWELL, S. & ALTMAN, D., 2011. REPORTING OF EFFECT DIRECTION AND SIZE IN ABSTRACTS OF SYSTEMATIC REVIEWS. JAMA, 306(18), PP. 1981-1982.

BENDER, R., 2009. INTRODUCTION TO THE USE OF REGRESSION MODELS IN EPIDEMIOLOGY. METHODS MOL BIOL, VOLUME 471, PP. 179-195.

BENDER, R. & BLETTNER, M., 2002. CALCULATING THE "NUMBER NEEDED TO BE EXPOSED" WITH ADJUSTMENT FOR CONFOUNDING VARIABLES IN EPIDEMIOLOGICAL STUDIES. J CLIN EPIDEMIOL, 55(5), PP. 525-530.

BENDER, R., KROMP, M., KIEFER, C. & STURTZ, S., 2013. ABSOLUTE RISKS RATHER THAN INCIDENCE RATES SHOULD BE USED TO ESTIMATE THE NUMBER NEEDED TO TREAT FROM TIME-TO-EVENT DATA. J CLIN EPIDEMIOL, 66(9), PP. 1038-1044.

BENDER, R., KROMP, M., KIEFER, C. & STURTZ, S., 2014. ESTIMATION OF NUMBERS NEEDED TO TREAT SHOULD BE BASED ON ABSOLUTE RISKS. J CLIN EPIDEMIOL, 67(2), PP. 238-239.

BENDER, R., KUSS, O., HILDEBRANDT, M. & GEHRMANN, U., 2007. ESTIMATING ADJUSTED NNT MEASURES IN LOGISTIC REGRESSION ANALYSIS. STAT MED, 26(30), PP. 5586-5595.

BERLIN, J. & GOLUB, R., 2014. META-ANALYSIS AS EVIDENCE: BUILDING A BETTER PYRAMID. JAMA, 312(6), PP. 603-606.

BERLIN, J., SOLEDAD-CEPEDA, M. & KIM, C., 2012. THE USE OF META-ANALYSIS IN PHARMACOEPIDEMIOLOGY. IN: B. STROM, S. KIMMEL & S. HENNESY, EDS. PHARMACOEPIDEMIOLOGY. 5TH ED. CHICHESTER, ENGLAND: JOHN WILEY & SONS, PP. 723-756.

BHATT, D. & MEHTA, C., 2016. ADAPTIVE DESIGNS FOR CLINICAL TRIALSV. N ENGL J MED, 375(1), PP. 65-74.

BIOMARKERS DEFINITIONS WORKING GROUP, 2001. BIOMARKERS AND SURROGATE ENDPOINTS: PREFERRED DEFINITIONS AND CONCEPTUAL FRAMEWORK. CLIN PHARMACOL THER, 69(3), PP. 89-95.

BJERRE, L. & LELORIER, J., 2000. EXPRESSING THE MAGNITUDE OF ADVERSE EFFECTS IN CASE-CONTROL STUDIES: "THE NUMBER OF PATIENTS NEEDED TO BE TREATED FOR ONE ADDITIONAL PATIENT TO BE HARMED". BMJ, 320(7233), PP. 503-506.

BORENSTEIN, M., HEDGES, L., HIGGINS, J. & ROTHSTEIN, H., 2009. INTRODUCTION TO META-ANALYSIS. CHICHESTER, ENGLAND: JOHN WILEY & SONS.

BORRACCI, R., PIÑEIRO, D. & ARRIBALZAGA, E., 2015. EFFECTS OF PRESENTING RISK INFORMATION IN DIFFERENT FORMATS TO CARDIOLOGISTS. A LATIN AMERICAN SURVEY. ARCH CARDIOL MEX, 85(1), PP. 3-8.

BOYD, C. ET AL., 2012. METHODS FOR BENEFIT AND HARM ASSESSMENT IN SYSTEMATIC REVIEWS. AHRQ PUBLICATION NO. 12(13)-EHC150-EF ED. ROCKVILLE: AGENCY FOR HEALTHCARE RESEARCH AND QUALITY.

BRATTON, D., PARMAR, M., PHILLIPS, P. & CHOODARI-OSKOOEI, B., 2016. TYPE I ERROR RATES OF MULTI-ARM MULTI-STAGE CLINICAL TRIALS: STRONG CONTROL AND IMPACT OF INTERMEDIATE OUTCOMES. TRIALS, 17(1), P. 309.

BROZEK, J. ET AL., 2009. GRADING QUALITY OF EVIDENCE AND STRENGTH OF RECOMMENDATIONS IN CLINICAL PRACTICE GUIDELINES. PART 1 OF 3. AN OVERVIEW OF THE GRADE APPROACH AND GRADING QUALITY OF EVIDENCE ABOUT INTERVENTIONS. ALLERGY, 64(5), PP. 669-677.

BUSSE, J. & GUYATT, G., 2015. COPRESENTATION OF RELATIVE AND ABSOLUTE EFFECTS IS ESSENTIAL TO PROMOTE OPTIMAL INTERPRETABILITY OF TREATMENT EFFECTS. J CLIN EPIDEMIOL, 68(3), PP. 355-356.

BYUN, J. ET AL., 2016. COMPARISON OF BENEFIT-RISK PREFERENCES OF PATIENTS AND PHYSICIANS REGARDING CYCLOOXYGENASE-2 INHIBITORS USING DISCRETE CHOICE EXPERIMENTS. PATIENT PREFER ADHERENCE, VOLUME 10, PP. 641-650.

CALFEE, J., 2006. A REPRESENTATIVE SURVEY OF M.S. PATIENTS ON ATTITUDES TOWARD THE BENEFITS AND RISKS OF DRUG THERAPY. WASHINGTON, DC: AEI-BROOKINGS JOINT CENTER FOR REGULATORY STUDIES.

CARPENTER, D., ZUCKER, E. & AVORN, J., 2008. DRUG-REVIEW DEADLINES AND SAFETY PROBLEMS. N ENGL J MED, 358(13), PP. 1354-1361.

CASARETT, D., 2016. THE SCIENCE OF CHOOSING WISELY - OVERCOMING THE THERAPEUTIC ILLUSION. N ENGL J MED, 374(13), PP. 1203-1205.

CATES, C., 2002. SIMPSON'S PARADOX AND CALCULATION OF NUMBER NEEDED TO TREAT FROM META-ANALYSIS. BMC MED RES METHODOL, 2(1).

CHALMERS, I. & ALTMAN, D., 1995. SYSTEMATIC REVIEW. LONDON, ENGLAND: BMJ PUBLISHING GROUP.

CHARLES, C., GAFNI, A. & WHELAN, T., 1997. SHARED DECISION-MAKING IN THE MEDICAL ENCOUNTER: WHAT DOES IT MEAN? (OR IT TAKES AT LEAST TWO TO TANGO). SOC SCI MED, 44(5), PP. 681-692.

CIOMS WORKING GROUP IV, 1998. BENEFIT-RISK BALANCE FOR MARKETED DRUGS: EVALUATING SAFETY SIGNALS, GENEVA: COUNCIL FOR INTERNATIONAL ORGANIZATIONS OF MEDICAL SCIENCES (CIOMS).

CITROME, L. & KETTER, T., 2013. WHEN DOES A DIFFERENCE MAKE A DIFFERENCE? INTERPRETATION OF NUMBER NEEDED TO TREAT, NUMBER NEEDED TO HARM, AND LIKELIHOOD TO BE HELPED OR HARMED. INT J CLIN PRACT, 67(5), PP. 407-411.

CLARK, L., FAIRHURST, C. & TORGERSON, D., 2016. ALLOCATION CONCEALMENT IN RANDOMISED CONTROLLED TRIALS: ARE WE GETTING BETTER? BMJ, VOLUME 355, P. I5663.

COLLET, D., 1994. MODELLING SURVIVAL DATA IN MEDICAL RESEARCH. CHAPMAN & HALL ED. LONDON: S.N.

COOK, R. & SACKETT, D., 1995. THE NUMBER NEEDED TO TREAT: A CLINICALLY USEFUL MEASURE OF TREATMENT EFFECT. BMJ, 310(6977), PP. 452-454.

COX, D. & SNELL, E., 1989. ANALYSIS OF BINARY DATA. LONDON: CHAPMAN AND HALL.

DA COSTA, B. ET AL., 2012. METHODS TO CONVERT CONTINUOUS OUTCOMES INTO ODDS RATIOS OF TREATMENT RESPONSE AND NUMBERS NEEDED TO TREAT: META-EPIDEMIOLOGICAL STUDY. INT J EPIDEMIOL, 41(5), PP. 1445-1459.

DAVIES, H., CROMBIE, I. & TAVAKOLI, M., 1998. WHEN CAN ODDS RATIOS MISLEAD? BMJ, 316(7136), PP. 989-991.

DEEKS, J., HIGGINS, J. & ALTMAN, D., 2011. CHAPTER 9: ANALYSING DATA AND UNDERTAKING META-ANALYSES. IN: T. C. COLLABORATION, ED. COCHRANE HANDBOOK FOR SYSTEMATIC REVIEWS OF INTERVENTIONS VERSION 5.1.0., S.L.: THE COCHRANE COLLABORATION.

EDWARDS, W., 1968. CONSERVATISM IN HUMAN INFORMATION PROCESSING. IN: FORMAL REPRESENTATIONS OF HUMAN JUDGMENT. NEW YORK: WILEY, PP. 17-52.

EDWARDS, W., PHILLIPS, L., HAYS, W. & GOODMAN, B., 1968. PROBABILISTIC INFORMATION PROCESSING SYSTEMS: DESIGN AND EVALUATION. IEEE TRANSACTIONS ON SYSTEMS SCIENCE AND CYBERNETICS, SSR-4, PP. 248-265.

EGAN, B. ET AL., 2016. 2013 ACC/AHA CHOLESTEROL GUIDELINE AND IMPLICATIONS FOR HEALTHY PEOPLE 2020 CARDIOVASCULAR DISEASE PREVENTION GOALS. J AM HEART ASSOC, 5(8), P. E003558.

EICHLER, H., ABADIE, E., BAKER, M. & RASI, G., 2012. FIFTY YEARS AFTER THALIDOMIDE; WHAT ROLE FOR DRUG REGULATORS? BR J CLIN PHARMACOL, 74(5), PP. 731-733.

EICHLER, H. ET AL., 2011. BRIDGING THE EFFICACY-EFFECTIVENESS GAP: A REGULATOR'S PERSPECTIVE ON ADDRESSING VARIABILITY OF DRUG RESPONSE. NAT REV DRUG DISCOV, 10(7), PP. 495-506.

EICHLER, H. ET AL., 2016. "THRESHOLD-CROSSING": A USEFUL WAY TO ESTABLISH THE COUNTERFACTUAL IN CLINICAL TRIALS? CLIN PHARMACOL THER, 100(6), PP. 699-712.

EICHLER, H. ET AL., 2013. THE RISKS OF RISK AVERSION IN DRUG REGULATION. NAT REV DRUG DISCOV, 12(12), PP. 907-916.

EICHLER, H. ET AL., 2008. BALANCING EARLY MARKET ACCESS TO NEW DRUGS WITH THE NEED FOR BENEFIT/RISK DATA: A MOUNTING DILEMMA. NAT REV DRUG DISCOV, 7(10), PP. 818-826.

EMA, 2005A. EUROPEAN MEDICINES AGENCY STATEMENT ON THE SUSPENSION OF USE OF BEXTRA [PRESS RELEASE]. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/EMA/INDEX.JSP?CURL=PAGES/NEWS_AND_EVENTS/NEWS/2009/11/NEWS_DETAIL_000237.JSP&MID=WC0B01AC0580 [ACCESSED 09 JAN 2017].

EMA, 2005B. EMEA PUBLIC STATEMENT ON THE SUSPENSION OF THE MARKETING AUTHORISATION FOR BEXTRA (VALDECOXIB) IN THE EUROPEAN UNION [PRESS RELEASE]. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/PUBLIC_STATEMENT/ [ACCESSED 09 JAN 2017].

EMA, 2007. COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE (CHMP). REFLECTION PAPER ON BENEFIT-RISK ASSESSMENT METHODS IN THE CONTEXT OF THE EVALUATION OF MARKETING AUTHORISATION APPLICATIONS OF MEDICINAL PRODUCTS FOR HUMAN USE (EMEA/CHMP/15404/2007). [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REGULATORY_AND_PROCEDURAL_GUIDELINE/2010/01/WC500069634.PDF [ACCESSED 10 MAY 2016].

EMA, 2008A. CHMP ASSESSMENT REPORT FOR TREDAPTIVE (DOC.REF.: EMEA/348364/2008). [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/EPAR_-_PUBLIC_ASSESSMENT_REPORT/HUMAN/000889/WC500042219.PDF [ACCESSED 09 FEBRUARY 2016].

EMA, 2009A. BENEFIT-RISK METHODOLOGY PROJECT. DEVELOPMENT AND TESTING OF TOOLS AND PROCESSES FOR BALANCING MULTIPLE BENEFITS AND RISKS AS AN AID TO INFORMED REGULATORY DECISIONS ABOUT MEDICINAL PRODUCTS.. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REPORT/2011/07/WC500109477.PDF [ACCESSED 15 FEB 2017].

EMA, 2010A. BENEFIT-RISK METHODOLOGY PROJECT: WORK PACKAGE 2 REPORT: APPLICABILITY OF CURRENT TOOLS AND PROCESSES FOR REGULATORY BENEFIT-RISK ASSESSMENT. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REPORT/2010/10/WC500097750.PDF [ACCESSED 15 FEB 2017].

EMA, 2011A. BENEFIT-RISK METHODOLOGY PROJECT - WORK PACKAGE 1 REPORT: DESCRIPTION OF THE CURRENT PRACTICE OF BENEFIT-RISK ASSESSMENT FOR CENTRALISED PROCEDURE PRODUCTS IN THE EU REGULATORY NETWORK. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REPORT/2011/07/WC500109478.PDF [ACCESSED 30 DEC 2016].

EMA, 2011B. BENEFIT–RISK METHODOLOGY PROJECT. WORK PACKAGE 3 REPORT: FIELD TESTS. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REPORT/2011/09/WC500112088.PDF [ACCESSED 29 DEC 2016].

EMA, 2012. BENEFIT-RISK METHODOLOGY PROJECT. REPORT ON RISK PERCEPTION STUDY MODULE. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REPORT/2012/02/WC500123226.PDF [ACCESSED 29 DEC 2016].

EMA, 2013A. MITIGATING RISK AVERSION IN MEDICINES REGULATION IN THE INTEREST OF PUBLIC HEALTH. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/EMA/INDEX.JSP?CURL=PAGES/NEWS_AND_EVENTS/NEWS/2013/11/NEWS_DETAIL_001957.JSP&MID=WC0B01AC058004D5C1 [ACCESSED 29 DEC 2016].

EMA, 2014A. GUIDELINE ON GOOD PHARMACOVIGILANCE PRACTICES (GVP) – MODULE V - RISK MANAGEMENT SYSTEMS (REV 1). [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/SCIENTIFIC_GUIDELINE/2012/06/WC500129134.PDF [ACCESSED 19 JAN 2017].

EMA, 2014B. BENEFIT-RISK METHODOLOGY PROJECT: UPDATE ON WORK PACKAGE 5: EFFECTS TABLE PILOT (PHASE I). [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REPORT/2014/02/WC500162036.PDF [ACCESSED 16 FEB 2017].

EMA, 2015A. GUIDANCE DOCUMENT FOR THE CONTENT OF THE <CO-> RAPPORTEUR DAY 80 CRITICAL ASSESSMENT REPORT: OVERVIEW AND LIST OF QUESTIONS. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REGULATORY_AND_PROCEDURAL_GUIDELINE/2009/10/WC500004800.PDF [ACCESSED 30 DEC 2016].

EMA, 2015B. ASSESSMENT TEMPLATES AND GUIDANCE. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/EMA/INDEX.JSP?CURL=PAGES/REGULATION/DOCUMENT_LISTING/DOCUMENT_LISTING_000337.JSP&MID=WC0B01AC0580022719 [ACCESSED 30 MAR 2017].

EMA, 2016A. EUROPEAN MEDICINES AGENCY - ABOUT US - WHAT WE DO. [ONLINE] AVAILABLE AT:

HTTP://WWW.EMA.EUROPA.EU/EMA/INDEX.JSP?CURL=PAGES/ABOUT_US/GENERAL/GENERAL_CONTENT_000091.JSP&MID=WC0B01AC0580028A42 [ACCESSED 27 DEC 2016].

EMA, 2016B. GUIDELINE ON THE SCIENTIFIC APPLICATION AND THE PRACTICAL ARRANGEMENTS NECESSARY TO IMPLEMENT REGULATION (EC) 507/2006 ON THE CONDITIONAL MARKETING AUTHORISATION FOR MEDICINAL PRODUCTS FOR HUMAN USE FALLING WITHIN THE SCOPE OF REGULATION (EC) 726/2004. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/SCIENTIFIC_GUIDELINE/2016/03/WC500202774.PDF [ACCESSED 10 JAN 2017].

EMA, 2016C. GUIDELINE ON GOOD PHARMACOVIGILANCE PRACTICES (GVP) - MODULE VIII – POST-AUTHORISATION SAFETY STUDIES (REV. 2). [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/SCIENTIFIC_GUIDELINE/2012/06/WC500129137.PDF [ACCESSED 19 JAN 2017].

ENGEL, P. ET AL., 2016. LESSONS LEARNED ON THE DESIGN AND THE CONDUCT OF POST-AUTHORIZATION SAFETY STUDIES: REVIEW OF 3 YEARS OF PRAC OVERSIGHT. BR J CLIN PHARMACOL.

ENGELS, E. ET AL., 2000. HETEROGENEITY AND STATISTICAL SIGNIFICANCE IN META-ANALYSIS: AN EMPIRICAL STUDY OF 125 META-ANALYSES. STAT MED, 19(13), PP. 1707-1728.

EUROPEAN UNION, 2001. DIRECTIVE 2001/83/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL OF 6 NOVEMBER 2001 ON THE COMMUNITY CODE RELATING TO MEDICINAL PRODUCTS FOR HUMAN USE.

EUROPEAN UNION, 2004. REGULATION NO. 726/2004 OF THE EP AND OF THE COUNCIL OF 31 MARCH 2004 LAYING DOWN COMMUNITY PROCEDURES FOR THE AUTHORISATION AND SUPERVISION OF MEDICINAL PRODUCTS FOR HUMAN AND VETERINARY USE AND ESTABLISHING A EUROPEAN MEDICINES AGENCY.

EVIDENCE-BASED MEDICINE WORKING GROUP, 1992. EVIDENCE-BASED MEDICINE. A NEW APPROACH TO TEACHING THE PRACTICE OF MEDICINE. JAMA, 268(17), PP. 2420-2425.

FALK, D. ET AL., 2014. CUMULATIVE PROPORTION OF RESPONDERS ANALYSIS (CPRA) AS A TOOL TO ASSESS TREATMENT OUTCOME IN ALCOHOL CLINICAL TRIALS. J STUD ALCOHOL DRUGS, 75(2), PP. 335-346.

FARRAR, J., DWORKIN, R. & MAX, M., 2006. USE OF THE CUMULATIVE PROPORTION OF RESPONDERS ANALYSIS GRAPH TO PRESENT PAIN DATA OVER A RANGE OF CUT-OFF POINTS: MAKING CLINICAL TRIAL DATA MORE UNDERSTANDABLE. J PAIN SYMPTOM MANAGE, 31(4), PP. 369-377.

FDA, 2005. FDA ISSUES PUBLIC HEALTH ADVISORY ON TYSABRI, A NEW DRUG FOR MS. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/NEWSEVENTS/NEWSROOM/PRESSANNOUNCEMENTS/2005/UCM108413.HTM [ACCESSED 29 DEC 2016].

FDA, 2013A. STRUCTURED APPROACH TO BENEFIT-RISK ASSESSMENT IN DRUG REGULATORY DECISION-MAKING: DRAFT PATIENT DRUG USER FREE ACT (PDUFA) I IMPLEMENTATION PLAN. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DOWNLOADS/FORINDUSTRY/USERFEES/PRESCRIPTIONDRUGUSERFEE/UCM329758.PDF [ACCESSED 02 APRIL 2015].

FDA, 2015. FAST TRACK, BREAKTHROUGH THERAPY, ACCELERATED APPROVAL, PRIORITY REVIEW. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/FORPATIENTS/APPROVALS/FAST/UCM20041766.HTM [ACCESSED 10 JAN 2017].

FLEMING, T. & DEMETS, D., 1996. SURROGATE END POINTS IN CLINICAL TRIALS: ARE WE BEING MISLED? ANN INTERN MED, 125(7), PP. 605-613.

FLEMING, T. & POWERS, J., 2012. BIOMARKERS AND SURROGATE ENDPOINTS IN CLINICAL TRIALS. STAT MED, 31(25), PP. 2973-2984.

FORROW, L., TAYLOR, W. & ARNOLD, R., 1992. ABSOLUTELY RELATIVE: HOW RESEARCH RESULTS ARE SUMMARIZED CAN AFFECT TREATMENT DECISIONS. AM J MED, 92(2), PP. 121-124.

FOWLER, F. J. ET AL., 2000. COMPARISON OF RECOMMENDATIONS BY UROLOGISTS AND RADIATION ONCOLOGISTS FOR TREATMENT OF CLINICALLY LOCALIZED PROSTATE CANCER. JAMA, 283(24), PP. 3217-3222.

FROUD, R., UNDERWOOD, M., CARNES, D. & ELDRIDGE, S., 2012. CLINICIANS' PERCEPTIONS OF REPORTING METHODS FOR BACK PAIN TRIALS: A QUALITATIVE STUDY. BR J GEN PRACT, 62(596), PP. E151-159.

FURUKAWA, T., GUYATT, G. & GRIFFITH, L., 2002. CAN WE INDIVIDUALIZE THE 'NUMBER NEEDED TO TREAT'? AN EMPIRICAL STUDY OF SUMMARY EFFECT MEASURES IN META-ANALYSES. INT J EPIDEMIOL, 31(1), PP. 72-76.

FURUKAWA, T. & LEUCHT, S., 2011. HOW TO OBTAIN NNT FROM COHEN'S D: COMPARISON OF TWO METHODS. PLOS ONE, 6(4), P. E19070.

GARATTINI, S. & BERTELE, V., 2007. HOW CAN WE REGULATE MEDICINES BETTER? BMJ, 335(7624), PP. 803-805.

GIGERENZER, G. ET AL., 2007. HELPING DOCTORS AND PATIENTS MAKE SENSE OF HEALTH STATISTICS. PSYCHOL SCI PUBLIC INTEREST, 8(2), PP. 53-96.

GIGERENZER, G., WEGWARTH, O. & FEUFEL, M., 2010. MISLEADING COMMUNICATION OF RISK. BMJ, VOLUME 341, P. C4830.

GRAHAM, J. & WIENER, J., 1995. RISK VS. RISK: TRADEOFFS IN PROTECTING HEALTH AND THE ENVIRONMENT. CAMBRIDGE, MASSACHUSETTS: HARVARD UNIVERSITY PRESS.

GRIMES, D. & SCHULZ, K., 2002A. BIAS AND CAUSAL ASSOCIATIONS IN OBSERVATIONAL RESEARCH. LANCET, 359(9302), PP. 248-252.

GRIMES, D. & SCHULZ, K., 2002B. COHORT STUDIES: MARCHING TOWARDS OUTCOMES. LANCET, 359(9303), PP. 341-345.

GROSS, C. ET AL., 2003. FACTORS INFLUENCING PHYSICIANS' REPORTED USE OF ANTICOAGULATION THERAPY IN NONVALVULAR ATRIAL FIBRILLATION: A CROSS-SECTIONAL SURVEY. CLIN THER, 25(6), PP. 1750-1764.

GRUNDMARK, B., HOLMBERG, L., GARMO, H. & ZETHELIUS, B., 2014. REDUCING THE NOISE IN SIGNAL DETECTION OF ADVERSE DRUG REACTIONS BY STANDARDIZING THE BACKGROUND: A PILOT STUDY ON ANALYSES OF PROPORTIONAL REPORTING RATIOS-BY-THERAPEUTIC AREA. EUR J CLIN PHARMACOL, 70(5), PP. 627-635.

GUO, J. ET AL., 2010. A REVIEW OF QUANTITATIVE RISK-BENEFIT METHODOLOGIES FOR ASSESSING DRUG SAFETY AND EFFICACY-REPORT OF THE ISPOR RISK-BENEFIT MANAGEMENT WORKING GROUP. VALUE HEALTH, 13(5), PP. 657-666.

GUYATT, G. ET AL., 2006. AN EMERGING CONSENSUS ON GRADING RECOMMENDATIONS? ACP J CLUB, 144(1), PP. A8-A9.

HALLGREEN, C. ET AL., 2014. BENEFIT-RISK ASSESSMENT IN A POST-MARKET SETTING: A CASE STUDY INTEGRATING REAL-LIFE EXPERIENCE INTO BENEFIT-RISK METHODOLOGY. PHARMACOEPIDEMIOL DRUG SAF, 23(9), PP. 974-983.

HAMMAD, T. ET AL., 2013. THE FUTURE OF POPULATION-BASED POSTMARKET DRUG RISK ASSESSMENT: A REGULATOR'S PERSPECTIVE. CLIN PHARMACOL THER, 94(3), PP. 349-358.

HAMMAD, T. ET AL., 2013. REPORTING OF META-ANALYSES OF RANDOMIZED CONTROLLED TRIALS WITH A FOCUS ON DRUG SAFETY: AN EMPIRICAL ASSESSMENT. CLIN TRIALS, 10(3), PP. 389-397.

HAMMAD, T., PINHEIRO, S. & NEYARAPALLY, G., 2011. SECONDARY USE OF RANDOMIZED CONTROLLED TRIALS TO EVALUATE DRUG SAFETY: A REVIEW OF METHODOLOGICAL CONSIDERATIONS. CLIN TRIALS, 8(5), PP. 559-570.

HAMMOND, J., KEENEY, R. & RAIFFA, H., 1999. SMART CHOICES: A PRACTICAL GUIDE TO MAKING BETTER DECISIONS. BOSTON, MA: HARVARD UNIVERSITY PRESS.

HASSELBLAD, V. & HEDGES, L., 1995. META-ANALYSIS OF SCREENING AND DIAGNOSTIC TESTS. PSYCHOL BULL, 117(1), PP. 167-178.

HAUKOOS, J. & LEWIS, R., 2015. THE PROPENSITY SCORE. JAMA, 314(15), PP. 1637-1638.

HAYNES, B. & HAINES, A., 1998. BARRIERS AND BRIDGES TO EVIDENCE BASED CLINICAL PRACTICE. BMJ, 317(7153), PP. 273-276.

HAZELL, L. & SHAKIR, S., 2006. UNDER-REPORTING OF ADVERSE DRUG REACTIONS: A SYSTEMATIC REVIEW. DRUG SAF, 29(5), PP. 385-396.

HELLER, R., DOBSON, A., ATTIA, J. & PAGE, J., 2002. IMPACT NUMBERS: MEASURES OF RISK FACTOR IMPACT ON THE WHOLE POPULATION FROM CASE-CONTROL AND COHORT STUDIES. J EPIDEMIOL COMMUNITY HEALTH, 56(8), PP. 606-610.

HENSCHKE, N., VAN ENST, A., FROUD, R. & OSTELO, R., 2014. RESPONDER ANALYSES IN RANDOMISED CONTROLLED TRIALS FOR CHRONIC LOW BACK PAIN: AN OVERVIEW OF CURRENTLY USED METHODS. EUR SPINE J, 23(4), PP. 772-778.

HIGGINS, J. & GREEN, S., 2011. COCHRANE HANDBOOK FOR SYSTEMATIC REVIEWS OF INTERVENTIONS VERSION 5.1.0., S.L. :THE COCHRANE COLLABORATION

HILDEBRANDT, M., VERVÖLGYI, E. & BENDER, R., 2009. CALCULATION OF NNTS IN RCTS WITH TIME-TO-EVENT OUTCOMES: A LITERATURE REVIEW. BMC MED RES METHODOL, 9(21).

HOEKMAN, J. ET AL., 2016. CHARACTERISTICS AND FOLLOW-UP OF POSTMARKETING STUDIES OF CONDITIONALLY AUTHORIZED MEDICINES IN THE EU. BR J CLIN PHARMACOL, 82(1), PP. 213-226.

HOFFMANN, T. & DEL MAR, C., 2017. CLINICIANS' EXPECTATIONS OF THE BENEFITS AND HARMS OF TREATMENTS, SCREENING, AND TESTS: A SYSTEMATIC REVIEW. JAMA INTERN MED, 177(3), PP. 407-419.

HUGHES, D. ET AL., 2016. RECOMMENDATIONS FOR BENEFIT-RISK ASSESSMENT METHODOLOGIES AND VISUAL REPRESENTATIONS. PHARMACOEPIDEMIOL DRUG SAF, 25(3), PP. 251-262.

ICH, 1998. THE INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE: ICH HARMONIZED TRIPARTITE GUIDELINE: STATISTICAL PRINCIPLES FOR CLINICAL TRIALS, E-9. [ONLINE] AVAILABLE AT: HTTP://WWW.ICH.ORG/FILEADMIN/PUBLIC_WEB_SITE/ICH_PRODUCTS/GUIDELINES/EFFICACY/E9/STEP4/E9_GUIDELINE.PDF [ACCESSED 02 JAN 2017].

JOHNSON, F. ET AL., 2010. ARE GASTROENTEROLOGISTS LESS TOLERANT OF TREATMENT RISKS THAN PATIENTS? BENEFIT-RISK PREFERENCES IN CROHN'S DISEASE MANAGEMENT. J MANAG CARE PHARM, 16(8), PP. 616-628.

JÜNI, P. ET AL., 2004. RISK OF CARDIOVASCULAR EVENTS AND ROFECOXIB: CUMULATIVE META-ANALYSIS. LANCET, 364(9450), PP. 2021-2029.

KAHNEMAN, D., 2002. MAPS OF BOUNDED RATIONALITY: A PERSPECTIVE ON INTUITIVE JUDGMENT AND CHOICE, S.L.: THE NOBEL FOUNDATION.

KANG, P., 2006. THE TYSABRI GAME PLAN. [ONLINE] AVAILABLE AT: HTTP://WWW.FORBES.COM/2006/03/08/TYSABRI-ELAN-BIOGEN-CX_PK_0308TYSABRI.HTML [ACCESSED 29 DEC 2016].

KATZ, N., PAILLARD, F. & VAN INWEGEN, R., 2015. A REVIEW OF THE USE OF THE NUMBER NEEDED TO TREAT TO EVALUATE THE EFFICACY OF ANALGESICS. J PAIN, 16(2), PP. 116-123.

KING, N., HARPER, S. & YOUNG, M., 2012. USE OF RELATIVE AND ABSOLUTE EFFECT MEASURES IN REPORTING HEALTH INEQUALITIES: STRUCTURED REVIEW. BMJ, VOLUME 345, P. E5774.

KLEIN, J. & MOESCHBERGER, M., 1997. SURVIVAL ANALYSIS: TECHNIQUES FOR CENSORED AND TRUNCATED DATA. NEW YORK: SPRINGER-VERLAG.

KRAEMER, H. & KUPFER, D., 2006. SIZE OF TREATMENT EFFECTS AND THEIR IMPORTANCE TO CLINICAL RESEARCH AND PRACTICE. BIOL PSYCHIATRY, 59(11), PP. 990-996.

LA ROCHELLE, P., LEXCHIN, J. & SIMONYAN, D., 2016. ANALYSIS OF THE DRUGS WITHDRAWN FROM THE US MARKET FROM 1976 TO 2010 FOR SAFETY REASONS. PHARM MED, 30(5), PP. 277-289.

LAINE, C., TAICHMAN, D. & MULROW, C., 2011. TRUSTWORTHY CLINICAL GUIDELINES. ANN INTERN MED, 154(11), PP. 774-775.

LARSSON, L. ET AL., 1996. THE SWEDISH RANDOMISED MAMMOGRAPHY SCREENING TRIALS: ANALYSIS OF THEIR EFFECT ON THE BREAST CANCER RELATED EXCESS MORTALITY. J MED SCREEN, 3(3), PP. 129-132.

LATHIA, C. ET AL., 2009. THE VALUE, QUALIFICATION, AND REGULATORY USE OF SURROGATE END POINTS IN DRUG DEVELOPMENT. CLIN PHARMACOL THER, 86(1), PP. 32-43.

LAUBENDER, R. & BENDER, R., 2010. ESTIMATING ADJUSTED RISK DIFFERENCE (RD) AND NUMBER NEEDED TO TREAT (NNT) MEASURES IN THE COX REGRESSION MODEL. STAT MED, 29(7-8), PP. 851-859.

LAUPACIS, A., SACKETT, D. & ROBERTS, R., 1988. AN ASSESSMENT OF CLINICALLY USEFUL MEASURES OF THE CONSEQUENCES OF TREATMENT. N ENGL J MED, 318(26), PP. 1728-1733.

LENERT, L., MARKOWITZ, D. & BLASCHKE, T., 1993. PRIMUM NON NOCERE? VALUING OF THE RISK OF DRUG TOXICITY IN THERAPEUTIC DECISION MAKING. CLIN PHARMACOL THER, 53(3), PP. 285-291.

LENERT, L., MARKOWITZ, D. & BLASCHKE, T., 1993. PRIMUM NON NOCERE? VALUING OF THE RISK OF DRUG TOXICITY IN THERAPEUTIC DECISION MAKING. CLIN PHARMACOL THER, 53(3), PP. 285-291.

LEONG, J., SALEK, S. & WALKER, S., 2015. BENEFIT-RISK ASSESSMENT OF MEDICINES: THE DEVELOPMENT AND APPLICATION OF A UNIVERSAL FRAMEWORK FOR DECISION-MAKING AND EFFECTIVE COMMUNCATION. SWITZERLAND: ADIS - SPRINGER INTERNATIONAL PUBLISHING.

LEYENS, L. ET AL., 2015. AVAILABLE TOOLS TO FACILITATE EARLY PATIENT ACCESS TO MEDICINES IN THE EU AND THE USA: ANALYSIS OF CONDITIONAL APPROVALS AND THE IMPLICATIONS FOR PERSONALIZED MEDICINE. PUBLIC HEALTH GENOMICS, 18(5), PP. 249-259.

LIBERTI, L. ET AL., 2015. OBSERVATIONS ON THREE ENDPOINT PROPERTIES AND THEIR RELATIONSHIP TO REGULATORY OUTCOMES OF EUROPEAN ONCOLOGY MARKETING APPLICATIONS. ONCOLOGIST, 20(6), PP. 683-691.

LIN, C., WU, S. & BALAKRISHNAN, N., 2003. PARAMETER ESTIMATION FOR THE LINEAR HAZARD RATE DISTRIBUTION BASED ON RECORDS AND INTER-RECORD TIMES. COMMUN STAT A, 32(4), PP. 729-748.

LUBSEN, J., HOES, A. & GROBBEE, D., 2000. IMPLICATIONS OF TRIAL RESULTS: THE POTENTIALLY MISLEADING NOTIONS OF NUMBER NEEDED TO TREAT AND AVERAGE DURATION OF LIFE GAINED. LANCET, 356(9243), PP. 1757-1759.

MA, H. ET AL., 2016. CONSIDERATIONS ON ENDPOINT SELECTION, WEIGHTING DETERMINATION, AND UNCERTAINTY EVALUATION IN THE BENEFIT-RISK ASSESSMENT OF MEDICAL PRODUCT. 8(4), PP. 417-425.

MANN, R. & ANDREWS, E., 2014. INTRODUCTION: UPDATED FROM SECOND EFITION. IN: E. ANDREWS & N. MOORE, EDS. PHARMACOVIGILANCE. 3RD ED. CHICHESTER: JOHN WILEY & SONS, LTD., PP. 1-9.

MARX, A. & BUCHER, H., 2003. NUMBERS NEEDED TO TREAT DERIVED FROM META-ANALYSIS: A WORD OF CAUTION. ACP J CLUB, 138(2), PP. A11-A12.

MAYNE, T., WHALEN, E. & VU, A., 2006. ANNUALIZED WAS FOUND BETTER THAN ABSOLUTE RISK REDUCTION IN THE CALCULATION OF NUMBER NEEDED TO TREAT IN CHRONIC CONDITIONS. J CLIN EPIDEMIOL, 59(3), PP. 217-223.

MCALISTER, F., 2002. COMMENTARY: RELATIVE TREATMENT EFFECTS ARE CONSISTENT ACROSS THE SPECTRUM OF UNDERLYING RISKS...USUALLY. INT J EPIDEMIOL, 31(1), PP. 76-77.

MCALISTER, F., 2008. THE "NUMBER NEEDED TO TREAT" TURNS 20 - AND CONTINUES TO BE USED AND MISUSED. CMAJ, 179(6), PP. 549-553.

MCALISTER, F., STRAUS, S., GUYATT, G. & HAYNES, R., 2000. USERS' GUIDES TO THE MEDICAL LITERATURE: XX. INTEGRATING RESEARCH EVIDENCE WITH THE CARE OF THE INDIVIDUAL PATIENT. EVIDENCE-BASED MEDICINE WORKING GROUP. JAMA, 283(21), PP. 2829-2836.

MCGETTIGAN, P. ET AL., 1999. THE EFFECTS OF INFORMATION FRAMING ON THE PRACTICES OF PHYSICIANS. J GEN INTERN MED, 14(10), PP. 633-642.

MCNAUGHTON, R., HUET, G. & SHAKIR, S., 2014. AN INVESTIGATION INTO DRUG PRODUCTS WITHDRAWN FROM THE EU MARKET BETWEEN 2002 AND 2011 FOR SAFETY REASONS AND THE EVIDENCE USED TO SUPPORT THE DECISION-MAKING. BMJ OPEN, 4(1), P. E004221.

MCQUAY, H. & MOORE, R., 1997. USING NUMERICAL RESULTS FROM SYSTEMATIC REVIEWS IN CLINICAL PRACTICE. ANN INTERN MED, 126(9), PP. 712-720.

MELLERS, B. & LOCKE, C., 2007. WHAT HAVE WE LEARNED FROM OUT MISTAKES? IN: ADVANCES IN DECISION ANALYSIS: FROM FOUNDATIONS TO APPLICATIONS. CAMBRIDGE: CAMBRIDGE UNIVERSITY PRESS, PP. 351-374.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2015. NUMBER NEEDED TO HARM IN THE POST-MARKETING SAFETY EVALUATION: RESULTS FOR ROSIGLITAZONE AND PIOGLITAZONE. PHARMACOEPIDEMIOL DRUG SAF, 24(12), PP. 1259-1270.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2016A. TESTING THE USEFULNESS OF THE NUMBER NEEDED TO TREAT TO BE HARMED (NNTH) IN BENEFIT-RISK EVALUATIONS: CASE STUDY WITH MEDICINES WITHDRAWN FROM THE EUROPEAN MARKET DUE TO SAFETY REASONS. EXPERT OPIN DRUG SAF, 15(10), PP. 1301-1312.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2016B. BENEFIT-RISK OF THERAPIES FOR RELAPSING-REMITTING MULTIPLE SCLEROSIS: TESTING THE NUMBER NEEDED TO TREAT TO BENEFIT (NNTB), NUMBER NEEDED TO TREAT TO HARM (NNTH) AND THE LIKELIHOOD TO BE HELPED OR HARMED (LHH): A SYSTEMATIC REVIEW AND META-ANALYSIS. CNS DRUGS, 30(10), PP. 909-929.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2017. NUMBER NEEDED TO TREAT (NNT) IN CLINICAL LITERATURE: AN APPRAISAL. BMC MEDICINE, DOI: 10.1186/S12916-017-0875-8.

MILLS, E., AYERS, D., CHOU, R. & THORLUND, K., 2015. ARE CURRENT STANDARDS OF REPORTING QUALITY FOR CLINICAL TRIALS SUFFICIENT IN ADDRESSING IMPORTANT SOURCES OF BIAS? CONTEMP CLIN TRIALS, 45(PT A), PP. 2-7.

MOL, P. ET AL., 2015. UNDERSTANDING DRUG PREFERENCES, DIFFERENT PERSPECTIVES. BR J CLIN PHARMACOL, 79(6), PP. 978-987.

MOORE, T. & FURBERG, C., 2012. THE SAFETY RISKS OF INNOVATION: THE FDA'S EXPEDITED DRUG DEVELOPMENT PATHWAY. JAMA, 308(9), PP. 869-870.

MOORE, R., MOORE, O., DERRY, S. & MCQUAY, H., 2008. NUMBERS NEEDED TO TREAT CALCULATED FROM RESPONDER RATES GIVE A BETTER INDICATION OF EFFICACY IN OSTEOARTHRITIS TRIALS THAN MEAN PAIN SCORES. ARTHRITIS RES THER, 10(2), P. R39.

MOORE, R. ET AL., 2010. RESPONDER ANALYSIS FOR PAIN RELIEF AND NUMBERS NEEDED TO TREAT IN A META-ANALYSIS OF ETORICOXIB OSTEOARTHRITIS TRIALS: BRIDGING A GAP BETWEEN CLINICAL TRIALS AND CLINICAL PRACTICE. ANN RHEUM DIS, 69(2), PP. 374-379.

MT-ISA, S. ET AL., 2014. BALANCING BENEFIT AND RISK OF MEDICINES: A SYSTEMATIC REVIEW AND CLASSIFICATION OF AVAILABLE METHODOLOGIES. PHARMACOEPIDEMIOL DRUG SAF, 23(7), PP. 667-678.

MURTHY, S., KAULDHER, S. & TARGOWNIK, L., 2006. PHYSICIANS' APPROACHES TO THE USE OF GASTROPROTECTIVE STRATEGIES IN LOW-RISK NON-STEROIDAL ANTI-INFLAMMATORY DRUG USERS. ALIMENT PHARMACOL THER, 23(9), PP. 1365-1372.

NAYLOR, C., CHEN, E. & STRAUSS, B., 1992. MEASURED ENTHUSIASM: DOES THE METHOD OF REPORTING TRIAL RESULTS ALTER PERCEPTIONS OF THERAPEUTIC EFFECTIVENESS?. ANN INTERN MED, 117(11), PP. 916-921.

NEXOE, J. ET AL., 2002. DANISH GPS' PERCEPTION OF DISEASE RISK AND BENEFIT OF PREVENTION. FAM PRACT, 19(1), PP. 3-6.

NISSEN, S. & WOLSKI, K., 2010. ROSIGLITAZONE REVISITED: AN UPDATED META-ANALYSIS OF RISK FOR MYOCARDIAL INFARCTION AND CARDIOVASCULAR MORTALITY. ARCH INTERN MED, 170(14), PP. 1191-1201.

NIXON, R. ET AL., 2016. A CASE STUDY USING THE PROACT-URL AND BRAT FRAMEWORKS FOR STRUCTURED BENEFIT RISK ASSESSMENT. BIOM J, 58(1), PP. 8-27.

ONAKPOYA, I., HENEGHAN, C. & ARONSON, J., 2015. DELAYS IN THE POST-MARKETING WITHDRAWAL OF DRUGS TO WHICH DEATHS HAVE BEEN ATTRIBUTED: A SYSTEMATIC INVESTIGATION AND ANALYSIS. BMC MED, ISSUE 13, P. 26.

ONAKPOYA, I., HENEGHAN, C. & ARONSON, J., 2016. POST-MARKETING WITHDRAWAL OF 462 MEDICINAL PRODUCTS BECAUSE OF ADVERSE DRUG REACTIONS: A SYSTEMATIC REVIEW OF THE WORLD LITERATURE. BMC MED, ISSUE 14, P. 10.

OSHIMA-LEE, E. & EMANUEL, E., 2013. SHARED DECISION MAKING TO IMPROVE CARE AND REDUCE COSTS. N ENGL J MED, 368(1), PP. 6-8.

OSIRI, M. ET AL., 2003. NUMBER NEEDED TO TREAT (NNT): IMPLICATION IN RHEUMATOLOGY CLINICAL PRACTICE. ANN RHEUM DIS, 62(4), PP. 316-321.

OZEN, G. ET AL., 2016. RISK OF DIABETES MELLITUS ASSOCIATED WITH DISEASE-MODIFYING ANTIRHEUMATIC DRUGS AND STATINS IN RHEUMATOID ARTHRITIS. ANN RHEUM DIS, PP. PII: ANNRHEUMDIS-2016-209954.

PALUDAN-MÜLLER, A., TEINDL LAURSEN, D. & HRÓBJARTSSON, A., 2016. MECHANISMS AND DIRECTION OF ALLOCATION BIAS IN RANDOMISED CLINICAL TRIALS. BMC MED RES METHODOL, 16(1), P. 133.

PALUDETTO, M., OLIVIER-ABBAL, P. & MONTASTRUC, J., 2012. IS SPONTANEOUS REPORTING ALWAYS THE MOST IMPORTANT INFORMATION SUPPORTING DRUG WITHDRAWALS FOR PHARMACOVIGILANCE REASONS IN FRANCE? PHARMACOEPIDEMIOL DRUG SAF, 21(12), PP. 1289-1294.

PATRICK, D. ET AL., 2011. CONTENT VALIDITY-ESTABLISHING AND REPORTING THE EVIDENCE IN NEWLY DEVELOPED PATIENT-REPORTED OUTCOMES (PRO) INSTRUMENTS FOR MEDICAL PRODUCT EVALUATION: ISPOR PRO GOOD RESEARCH PRACTICES TASK FORCE REPORT: PART 1-ELICITING CONCEPTS FOR A NEW PRO INSTRUMENT. VALUE HEALTH, 14(8), PP. 967-977.

PERNEGER, T. & AGORITSAS, T., 2011. DOCTORS AND PATIENTS' SUSCEPTIBILITY TO FRAMING BIAS: A RANDOMIZED TRIAL. J GEN INTERN MED, 26(12), PP. 1411-1417.

PIGNATTI, F. ET AL., 2015. STRUCTURED FRAMEWORKS TO INCREASE THE TRANSPARENCY OF THE ASSESSMENT OF BENEFITS AND RISKS OF MEDICINES: CURRENT STATUS AND POSSIBLE FUTURE DIRECTIONS. CLIN PHARMACOL THER, 98(5), PP. 522-533.

PROTECT, 2013. REVIEW OF METHODOLOGIES FOR BENEFIT AND RISK ASSESSMENT OF MEDICATION. [ONLINE] AVAILABLE AT: HTTP://WWW.IMI-PROTECT.EU/DOCUMENTS/SHAHRULETALREVIEWOFMETHODOLOGIESFORBENEFITANDRISKASSESSMENTOFMEDICATIONMAY2013.PDF [ACCESSED 29 DEC 2016].

PROTECT, 2011. ABOUT PROTECT. [ONLINE] AVAILABLE AT: HTTP://WWW.IMI-PROTECT.EU/ABOUT.SHTML [ACCESSED 16 FEB 2017].

PROTECT, 2011. WP5: BENEFIT-RISK INTEGRATION AND REPRESENTATION. [ONLINE] AVAILABLE AT: HTTP://WWW.IMI-PROTECT.EU/WP5.SHTML [ACCESSED 16 FEB 2017].

RAWLINS, M., 1987. RISK-BENEFIT DECISIONS IN LICENSING CHANGES. IN: MEDICINES AND RISK/BENEFIT DECISIONS. LANCASTER: MTP PRESS LIMITED, PP. 137-141.

RAWLINS, M., 2004. CUTTING THE COST OF DRUG DEVELOPMENT? NAT REV DRUG DISCOV, 3(4), PP. 360-364.

RAWLINS, M. & THOMPSON, J., 1977. PATHOGENESIS OF ADVERSE DRUG REACTIONS. IN: TEXTBOOK OF ADVERSE DRUG REACTIONS. OXFORD: OXFORD UNIVERSITY PRESS.

ROSENBERG, L., COOGAN, P. & PALMER, J., 2012. CASE-CONTROL SURVEILLANCE. IN: B. STROM, S. KIMMEL & S. HENNESSY, EDS. PHARMACOEPIDEMIOLOGY. 5TH ED. CHICHESTER, ENGLAND: JOHN WILEY & SONS, PP. 287-300.

ROTHMAN, K., 2010. CURBING TYPE I AND TYPE II ERRORS. EUR J EPIDEMIOL, 25(4), PP. 223-224.

RUBIN, D., 2007. THE DESIGN VERSUS THE ANALYSIS OF OBSERVATIONAL STUDIES FOR CAUSAL EFFECTS: PARALLELS WITH THE DESIGN OF RANDOMIZED TRIALS. STAT MED, 26(1), PP. 20-36.

SACKETT, D. ET AL., 1996. EVIDENCE BASED MEDICINE: WHAT IT IS AND WHAT IT ISN'T. BMJ, 312(7023), PP. 71-72.

SACKETT, D., 2001. WHY RANDOMIZED CONTROLLED TRIALS FAIL BUT NEEDN'T: 2. FAILURE TO EMPLOY PHYSIOLOGICAL STATISTICS, OR THE ONLY FORMULA A CLINICIAN-TRIALIST IS EVER LIKELY TO NEED (OR UNDERSTAND!). CMAJ, 165(9), PP. 1226-1237.

SACKS, L. ET AL., 2014. SCIENTIFIC AND REGULATORY REASONS FOR DELAY AND DENIAL OF FDA APPROVAL OF INITIAL APPLICATIONS FOR NEW DRUGS, 2000-2012. JAMA, 311(4), PP. 378-384.

SAVOVIĆ, J. ET AL., 2012. INFLUENCE OF REPORTED STUDY DESIGN CHARACTERISTICS ON INTERVENTION EFFECT ESTIMATES FROM RANDOMIZED, CONTROLLED TRIALS. ANN INTERN MED, 157(6), PP. 429-438.

SCHMID, C., LAU, J., MCINTOSH, M. & CAPPELLERI, J., 1998. AN EMPIRICAL STUDY OF THE EFFECT OF THE CONTROL RATE AS A PREDICTOR OF TREATMENT EFFICACY IN META-ANALYSIS OF CLINICAL TRIALS. STAT MED, 17(17), PP. 1923-1942.

SCHULTZ, W., 2007. BOLSTERING THE FDA'S DRUG-SAFETY AUTHORITY. N ENGL J MED, 357(22), PP. 2217-2219.

SCHÜNEMANN, H. ET AL., 2011. CHAPTER 12: INTERPRETING RESULTS AND DRAWING CONCLUSIONS. IN: J. HIGGINS & S. GREEN, EDS. COCHRANE HANDBOOK FOR SYSTEMATIC REVIEWS OF INTERVENTIONS VERSION 5.1.0., S.L.: THE COCHRANE COLLABORATION.

SCHWARTZ, L., WOLOSHIN, S., DVORIN, E. & WELCH, H., 2006. RATIO MEASURES IN LEADING MEDICAL JOURNALS: STRUCTURED REVIEW OF ACCESSIBILITY OF UNDERLYING ABSOLUTE RISKS. BMJ, 333(7581), P. 1248.

SEDRAKYAN, A. & SHIH, C., 2007. IMPROVING DEPICTION OF BENEFITS AND HARMS: ANALYSES OF STUDIES OF WELL-KNOWN THERAPEUTICS AND REVIEW OF HIGH-IMPACT MEDICAL JOURNALS. MED CARE, 45(10 SUPPL 2), PP. S23-28.

SEVER, P., POULTER, N., DAHLÖF, B. & WEDEL, H., 2005. DIFFERENT TIME COURSE FOR PREVENTION OF CORONARY AND STROKE EVENTS BY ATORVASTATIN IN THE ANGLO-SCANDINAVIAN CARDIAC OUTCOMES TRIAL-LIPID-LOWERING ARM (ASCOT-LLA). AM J CARDIOL, 56(5A), PP. 39F-44F.

SLOVIC, P., FINUCANE, M., PETERS, E. & MACGREGOR, D., 2004. RISK AS ANALYSIS AND RISK AS FEELINGS: SOME THOUGHTS ABOUT AFFECT, REASON, RISK, AND RATIONALITY. RISK ANAL, 24(2), PP. 311-322.

SMEETH, L., HAINES, A. & EBRAHIM, S., 1999. NUMBERS NEEDED TO TREAT DERIVED FROM META-ANALYSES–SOMETIMES INFORMATIVE, USUALLY MISLEADING. BMJ, 318(7197), PP. 1548-1551.

STANG, A., POOLE, C. & BENDER, R., 2010. COMMON PROBLEMS RELATED TO THE USE OF NUMBER NEEDED TO TREAT. J CLIN EPIDEMIOL, 63(8), PP. 820-825.

STIGGELBOUT, A. ET AL., 2012. SHARED DECISION MAKING: REALLY PUTTING PATIENTS AT THE CENTRE OF HEALTHCARE. BMJ, VOLUME 344, P. E256.

STOVITZ, S. & SHRIER, I., 2013. MEDICAL DECISION MAKING AND THE IMPORTANCE OF BASELINE RISK. BR J GEN PRACT, 63(616), PP. E795-797.

STRAUS, S., GLASZIOU, P., RICHARDSON, W. & HAYNES, R., 2011. EVIDENCE-BASED MEDICINE: HOW TO PRACTICE AND TEACH IT. 4TH ED. LONDON, UK: CHURCHILL LIVINGSTONE.

STROM, B., 2012. BASIC PRINCIPLES OF CLINICAL EPIDEMIOLOGY RELEVANT TO PHARMACOEPIDEMIOLOGIC STUDIES. IN: B. STROM, S. KIMMEL & S. HENNESSY, EDS. PHARMACOEPIDEMIOLOGY. 5TH ED. CHICHESTER, ENGLAND: JOHN WILEY & SONS, PP. 38-51.

STROM, B., 2013. STUDY DESIGNS AVAILABLE FOR PHARMACOEPIDEMIOLOGY STUDIES. IN: B. STROM, S. KIMMEL & S. HENNESSY, EDS. TEXTBOOK OF PHARMACOEPIDEMIOLOGY. CHICHESTER, ENGLAND: JOHN WILEY & SONS, PP. 17-29.

SUISSA, D., BRASSARD, P., SMIECHOWSKI, B. & SUISSA, S., 2012. NUMBER NEEDED TO TREAT IS INCORRECT WITHOUT PROPER TIME-RELATED CONSIDERATIONS. J CLIN EPIDEMIOL, 65(1), PP. 42-46.

SUISSA, S., 1991. BINARY METHODS FOR CONTINUOUS OUTCOMES: A PARAMETRIC ALTERNATIVE. J CLIN EPIDEMIOL, 44(3), PP. 241-248.

SUISSA, S., 2015. THE NUMBER NEEDED TO TREAT: 25 YEARS OF TRIALS AND TRIBULATIONS IN CLINICAL RESEARCH. RAMBAM MAIMONIDES MED J, 6(3).

TEMPLE, R., 1995. A REGULATORY AUTHORITY'S OPINION ABOUT SURROGATE ENDPOINTS. IN: CLINICAL MEASUREMENT IN DRUG EVALUATION. NEW YORK: J. WILEY.

THE MEDICINES IN EUROPE FORUM, 2007. HAI EUROPE & THE ISDB. OPEN LETTER TO EC PRESIDENT BARROSO. [ONLINE] AVAILABLE AT: HTTP://WWW.ISDBWEB.ORG/DOCUMENTS/UPLOADS/OPENLETTER.PDF [ACCESSED 09 JAN 2017].

THORLUND, K. ET AL., 2011. POOLING HEALTH-RELATED QUALITY OF LIFE OUTCOMES IN META-ANALYSIS-A TUTORIAL AND REVIEW OF METHODS FOR ENHANCING INTERPRETABILITY. RES SYNTH METHODS, 2(3), PP. 188-203.

TROJANO, M. ET AL., 2017. TREATMENT DECISIONS IN MULTIPLE SCLEROSIS - INSIGHTS FROM REAL-WORLD OBSERVATIONAL STUDIES. NAT REV NEUROL, 13(2), PP. 105-118.

UETRECHT, J. & NAISBITT, D., 2013. IDIOSYNCRATIC ADVERSE DRUG REACTIONS: CURRENT CONCEPTS. PHARMACOL REV, 65(2), PP. 779-808.

VAN STAA, T. ET AL., 2008. EVALUATING DRUG TOXICITY SIGNALS: IS A HIERARCHICAL CLASSIFICATION OF EVIDENCE USEFUL OR A HINDRANCE? PHARMACOEPIDEMIOL DRUG SAF, 17(5), PP. 475-484.

VERMEER, N. ET AL., 2014. RISK MANAGEMENT PLANS AS A TOOL FOR PROACTIVE PHARMACOVIGILANCE: A COHORT STUDY OF NEWLY APPROVED DRUGS IN EUROPE. CLIN PHARMACOL THER, 96(6), PP. 723-731.

WALKER, S. ET AL., 2015. A UNIVERSAL FRAMEWORK FOR THE BENEFIT-RISK ASSESSMENT OF MEDICINES: IS THIS THE WAY FORWARD? THERAPEUTIC INNOVATION REGULATORY SCI, 49(1), PP. 17-25.

WEINTRAUB, W., LÜSCHER, T. & POCOCK, S., 2015. THE PERILS OF SURROGATE ENDPOINTS. EUR HEART J, 36(33), PP. 2212-2218.

XIE, J. & LIU, C., 2005. ADJUSTED KAPLAN-MEIER ESTIMATOR AND LOG-RANK TEST WITH INVERSE PROBABILITY OF TREATMENT WEIGHTING FOR SURVIVAL DATA. STAT MED, 24(20), PP. 3089-3110.

YUAN, Z., LEVITAN, B. & BERLIN, J., 2011. BENEFIT-RISK ASSESSMENT: TO QUANTIFY OR NOT TO QUANTIFY, THAT IS THE QUESTION. PHARMACOEPIDEMIOL DRUG SAF, 20(6), PP. 653-656.

YU, T. ET AL., 2015. USE OF SURROGATE OUTCOMES IN US FDA DRUG APPROVALS, 2003-2012: A SURVEY. BMJ OPEN, 5(11), P. E007960.

ZHANG, J. & YU, K., 1998. WHAT'S THE RELATIVE RISK? A METHOD OF CORRECTING THE ODDS RATIO IN COHORT STUDIES OF COMMON OUTCOMES. JAMA, 280(19), PP. 1690-1691.

CHAPTER II – NUMBER NEEDED TO HARM IN THE POST-MARKETING SAFETY EVALUATION: RESULTS FOR ROSIGLITAZONE AND PIOGLITAZONE

II. NUMBER NEEDED TO HARM IN THE POST-MARKETING SAFETY EVALUATION: RESULTS FOR ROSIGLITAZONE AND PIOGLITAZONE

II.1. ABSTRACT

Our aim was to investigate the usefulness of metric indices in post-marketing safety evaluations by estimating number needed to harm (NNTH) values for cardiovascular (CV) adverse outcomes for rosiglitazone and pioglitazone. Reports from regulatory authorities (RAs) were consulted, and Medline searches were performed to identify studies assessing CV risks [all-cause death, CV death, myocardial infarction (MI), stroke, or congestive heart failure (CHF)] for thiazolidinediones. Meta-analyses were performed to pool evidence from randomized controlled trials (RCTs) and observational studies. NNTHs [with 95% confidence intervals (CI)] per year were estimated for CV adverse events. Reports from RAs included two meta-analyses of short-term RCTs, two long-term RCTs (RECORD and PROACTIVE), and a systematic review of observational studies (n= 29). The Medline search identified six additional observational studies. Statistically significant NNTH values were obtained for the following: (i) rosiglitazone versus control on MI and CHF in the meta-analysis of RCTs (NNTH 16, 95% CI 10–255; and NNTH 7; 95% CI 5–16, respectively) and meta-analysis of observational studies (NNTH 12, 95% CI 9–20; and NNTH 5, 95% CI 32–131, respectively) and on CHF in the RECORD (NNTH 6, 95% CI 4–14); (ii) pioglitazone versus control on CHF (NNTH 11, 95% CI 6–403) in the meta-analysis of RCTs and PROACTIVE (NNTH 12, 95% CI 8–43); and (iii) rosiglitazone versus pioglitazone on MI (NNTH 69, 95% CI 32–379), stroke (NNTH 36, 95% CI 20–225), CHF (NNTH 33, 95% CI 19–47), and all-cause death (NNTH 63, 95% CI 49–100) in the meta-analysis of observational studies. The NNTH values suggested an increased CV risk with rosiglitazone versus pioglitazone across several sources of information. The inclusion of objective metrics in post-marketing drug's benefit–risk assessments could be of increased value and help RAs to make consistent decisions on drug safety.

II.2. INTRODUCTION

Several drugs have been withdrawn from the market because of safety reasons (Wysowski & Swartz 2005), (Clarke, Deeks & Shakir 2006), (Qureshi et al. 2011), (McNaughton, Huet & Shakir 2014). The decision of withdrawing a drug from the market has a major impact in the society and should be based on the best evidence available on benefits and harms (Clarke, Deeks & Shakir 2006), (Vandenbroucke & Psaty 2008), (Hammad et al. 2013). Because safety signals can arise from spontaneous reports of adverse events, observational studies, randomized controlled trials (RCTs), or meta-analyses (Lester et al. 2013), post-marketing benefit–risk assessment should consider data from all sources of evidence (Vandenbroucke & Psaty 2008), (AHRQ 2014). Methodologies used to integrate data from multiple sources have been discussed (Hammad et al. 2013), (Alves, Batel-Marques & Macedo 2012).

There is an increased interest from all stakeholders deciding on drug therapy in applying structured approaches for benefit–risk assessment that can bring clarity to the decision-making process and help ensure that different regulatory authorities make consistent decisions (Hammad et al. 2013), (Nixon et al. 2016), (EMA 2007), (FDA 2013a). The incorporation of quantitative methodologies into the process has been advocated as a contribution to improve regulatory decisions (Nixon et al. 2016), (Yuan, Levitan & Berlin 2011). However, the application of such methodologies for benefit–risk assessment remains elusive (Guo et al. 2010), (Eichler et al. 2013). In this context, the European Medicines Agency (EMA) initiated the PROTECT project, which is aimed to develop and test tools and processes for balancing benefits and risks, which could be used as an aid to make informed, science-based regulatory decisions (EMA 2011b), (Mt-Isa et al. 2013). Number needed to treat (NNT) is among the methodologies that were recommended for further examinations in benefit–risk assessment of drugs (Mt-Isa et al. 2014). NNT to benefit (NNTB) (or NNT to harm [NNTH]) is a measure of effect size that is defined as the number of patients who need to be treated with one therapy versus another in order to encounter an additional beneficial (or harmful) outcome of interest over a defined period (Laupacis, Sackett & Roberts 1988), (Cook & Sackett 1995), (Citrome & Ketter 2013).

Regulatory authorities may make different decisions despite having access to the same data (Walker et al. 2015). The benefit-risk ratio of rosiglitazone, a thiazolidinedione (TZD) used to treat type 2 diabetes mellitus, was re-assessed by regulatory authorities because of cardiovascular (CV) safety reasons. While the EMA decided to withdraw

rosiglitazone from the market, the USA Food and Drug Administration (FDA) left the drug in the market, although it has imposed some restrictions (EMA 2010b), (FDA 2011). Despite that both regulatory authorities have analysed the same data, they made divergent decisions. There is a rationale to investigate safety assessments of marketed drugs in the context of benefit–risk re-evaluations.

This study is aimed to investigate the usefulness of objective metric indices in post-marketing safety assessments through the estimation of NNTH values for CV adverse events associated with the use of rosiglitazone and pioglitazone.

II.3. METHODS

II.3.1. DATA SOURCES

A review of the published evidence was carried out to identify studies aimed to assess the risk of CV adverse events associated with the use of rosiglitazone or pioglitazone in patients with type 2 diabetes mellitus. Briefing documents from the FDA Division of Metabolism and Endocrinology Products Advisory Committee meetings and European Public Assessment Reports were retrieved from the FDA and the EMA websites, respectively.

Medline searches (up to 28 February 2015) were performed in order to update the evidence contained in documents produced by the regulatory authorities at the time of the re-assessment of the CV safety of rosiglitazone. First, we searched for RCTs designed to assess CV adverse events in association with the use of rosiglitazone or pioglitazone. A second Medline search was performed to identify observational studies designed to assess the CV risk of rosiglitazone or pioglitazone. The search strategies are described in Supplemental Table II. 1.

II.3.2. DATA EXTRACTION

Two reviewers (DMM and CCA) independently extracted data from the included studies. Discrepancies were resolved by majority (two out of three) decision involving a third investigator (FBM). Data on study characteristics (methodology, included population, study design, and drugs evaluated) and outcomes (CV adverse events) during treatment were extracted.

II.3.2.1. Outcomes assessed

The outcomes assessed were individual cases of all cause death, CV death, myocardial infarction (MI), stroke, and congestive heart failure (CHF). The definition of each event is provided elsewhere (FDA 2010a), (FDA 2013b).

II.3.3. STATISTICAL ANALYSIS

II.3.3.1. Set of analyses

Separate analyses were performed for rosiglitazone and pioglitazone based on several subgroups of studies. According to the experts from the FDA, carrying out separate analyses based on different subgroups of studies [i.e., separated according to the control group (placebo or active therapy) and to the regimen (monotherapy or add-on therapy)] allows for a better comparison between rosiglitazone and pioglitazone by eliminating some systematic differences between studies (FDA 2010a), (FDA 2013b). The first set of analyses was carried out by comparing TZD treatment with non-TZD antidiabetic treatment or placebo (overall results). The second set of analyses was carried out based on the control group, that is, TZD versus placebo control and TZD versus active control (non-TZD antidiabetic therapy). The third set of analyses comprised the following: (i) monotherapy studies (TZD monotherapy versus non-TZD antidiabetic monotherapy) and (ii) add-on studies (TZD added on to background therapy versus non-TZD antidiabetic added on to background therapy). A last set of analyses included studies that directly compared rosiglitazone with pioglitazone.

II.3.3.2. Analytic techniques

Meta-analyses were performed to determine pooled evidence from RCTs and observational studies whenever possible. Studies must have provided risk estimates [relative risk (RR), odds ratio (OR), or hazard ratio (HR)] for patients treated with rosiglitazone or pioglitazone compared with a control group. Because the CV adverse events assessed in this study can be considered as rare, similarity was assumed between RR, OR, or HR (Loke, Kwok & Singh 2011). The most adjusted estimate was used for studies presenting more than one risk estimate. Meta-analyses were conducted using a random-effects model in order to pool the OR with their 95% confidence intervals (CI) (DerSimonian & Laird 1986). This model was chosen as it is more conservative than a fixed-effect model in the presence of between-studies heterogeneity. Between studies heterogeneity was assessed using the I^2 measure of inconsistency (Higgins et al. 2003). All statistical analyses were performed using the COMPREHENSIVE META-ANALYSIS version 2 (Biostat, Englewood, NJ, USA).

II.3.3.3. Number needed to harm

Usually, NNTH is calculated by taking the reciprocal of the absolute risk increase between two groups when one is appraising dichotomous data from a single-study report (Straus et al. 2011). A different methodology was used in the present study because data were obtained from meta-analyses. NNTH per year (and 95% CI) was estimated for each CV adverse event by applying the pooled OR (and the limits of its corresponding 95% CI) from the meta-analyses (or individual studies when applicable) to baseline event rates per year (Straus et al. 2011). Baseline event rates per year for CV events were obtained from the Look AHEAD (Action for Health in Diabetes) Research Group trial (all-cause death, 0.86%; CV death, 0.24%; MI, 0.84%; stroke, 0.34%; and CHF, 0.51%) (Look AHEAD Research Group 2013). The following formula was used: NNTH = 1 + [(baseline event rate) × (OR – 1)] / [(1 - baseline event rate) × (baseline event rate) × (OR – 1)] (Straus et al. 2011). In case that the 95% CI for the NNTH estimate contain infinity, NNTH estimate is not statistically significant at the *p* threshold of <0.05.[2]

[2] In such cases, one of the confidence limits indicates harm (NNTH) and the other indicates benefit (NNTB), with the scale for NNT going from NNTH = 1 to NNTB = 1 via infinity. A negative value of NNTH is a positive value of NNTB (Altman, 1998).

II.4. RESULTS

The flow diagram of the study is presented in Figure II. 1. Documents from regulatory authorities included two meta-analyses of double-blind short-term RCTs (between 2 months and 2 years in duration) completed by December 2009 (rosiglitazone versus control, n= 1, and pioglitazone versus control, n=1), two long-term RCTs (>2 years in duration) (RECORD and PROACTIVE trials), and a systematic review of observational studies on both TZDs (seven nested case–control and 22 cohort studies) (FDA 2010a), (FDA 2013b), (EMA 2010c). The Medline search identified no further RCTs designed to assess CV safety of TZDs, but led to the identification of six additional observational studies. The characteristics of the observational studies are described in Supplemental Table II. 2. The results are presented in Table II. 1 (rosiglitazone versus non-TZD comparators), Table II. 2 (pioglitazone versus non-TZD comparators), and Table II. 3 (rosiglitazone versus pioglitazone). A more detailed description of results obtained in meta-analyses of observational studies is provided in Supplemental Table II. 3, Supplemental Table II. 4, and Supplemental Table II. 5.

II.4.1. ALL-CAUSE DEATH

According to the results of the meta-analyses of RCTs, rosiglitazone was not associated with an increased risk of all-cause death. When compared with placebo, pioglitazone reduced the risk of all-cause death (NNTB 19, 95% CI NNTB 9–408). Meta-analyses of observational studies produced conflicting results for rosiglitazone when used as add-on therapy versus non-TZD comparators depending on studies' design (case–control study, NNTH 40, 95% CI 27–77; cohort study, NNTB 68, 95% CI 35–408). However, when directly compared with pioglitazone, the use of rosiglitazone resulted in an increased risk of all-cause death (NNTH 63, 95% CI 49–100).

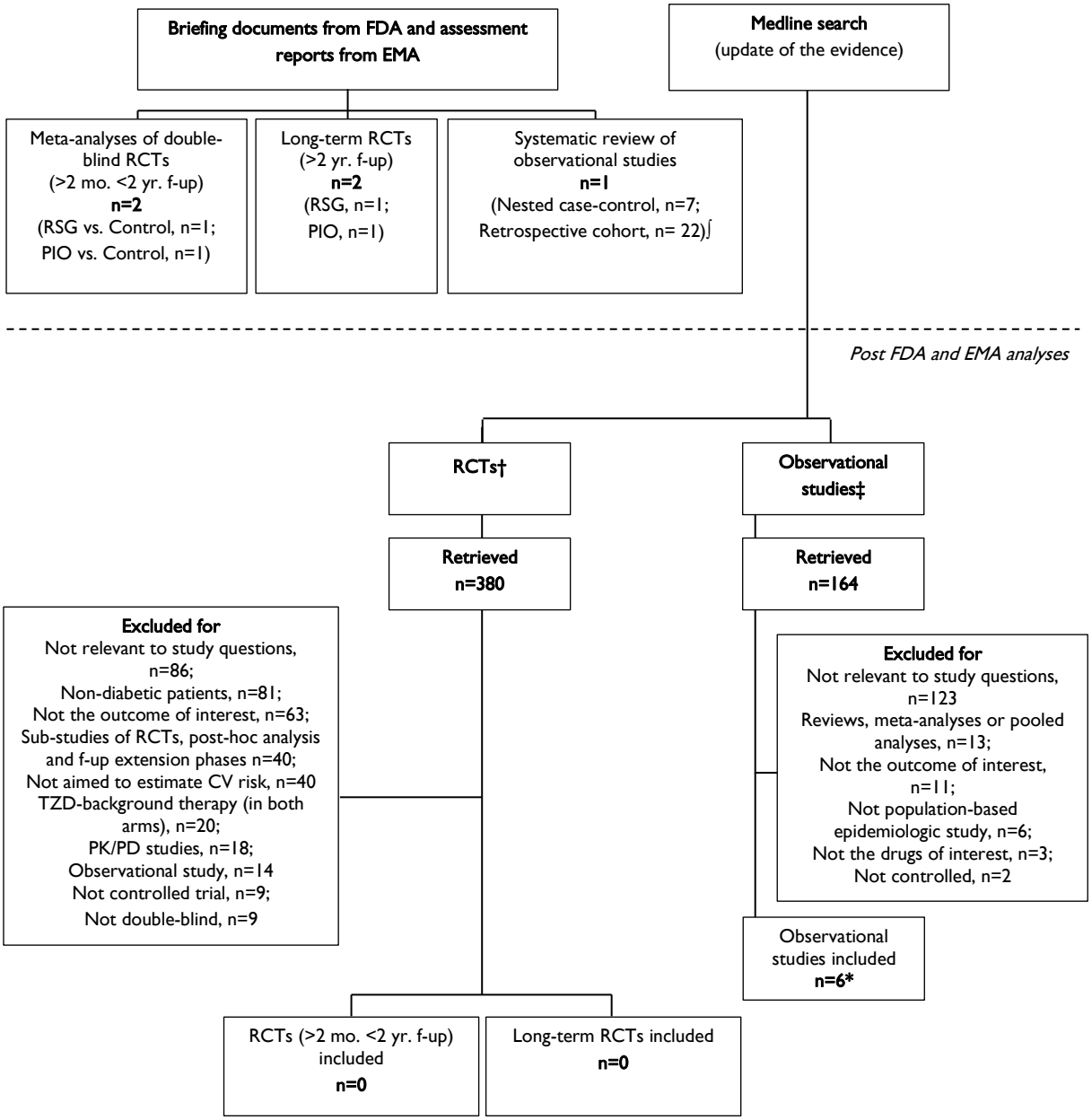


Figure II. 1 – Flow diagram of the available evidence for inclusion in the study.

FDA, Food and Drug Administration; EMA, European Medicines Agency; mo., months; PIO, pioglitazone; RCT, Randomized controlled trial; RSG, rosiglitazone; yr., years.

∫We have considered two additional retrospective cohort studies (Pantalone et al. 2009 and Ramirez et al. 2009) that FDA collaborators had excluded from their analysis, but we have not included a retrospective cohort study (Shaya et al. 2009) because it only reported results on a composite of cardiovascular events.

†Literature search from December 31, 2009 until February 28, 2015.

‡ Literature search from December 31, 2009 until February 28, 2015.

*A meta-analysis of observational studies was carried out by including the 6 studies found in the Medline search plus the 29 studies previously identified in the briefing documents from FDA.

Table II. 1 – Adjusted odds ratio (95% confidence intervals) and number needed to treat to harm (95% confidence intervals) for cardiovascular adverse events associated with the use of rosiglitazone in several settings according to each type of evidence.

Studies with rosiglitazone		FDA Meta-analysis of RCTs	RECORD Trial - original	RECORD Trial – re-adjudication	Meta-analysis of observational studies		
					Case-control	Cohort	Overall
Studies controlled with placebo							
MI	OR	**2.23 (1.14, 4.64)**	NA	NA	1.14 (0.90, 1.44)	0.86 (0.64, 1.17)	1.01 (0.77, 1.32)
	NNTH	**13 (9, 60)**	NA	NA	NNTH 59 (NNTH 24 to ∞ to NNTB 68)	NNTB 46 (NNTB 14 to ∞ to NNTH 51)	NNTH 751 (NNTH 30 to ∞ to NNTB 26)
Stroke	OR	0.65 (0.27, 1.52)	NA	NA	NA	NA	NA
	NNTH	NNTB 11 (NNTB 4 to ∞ to NNTH 11)	NA	NA	NA	NA	NA
CHF	OR	**2.20 (1.40, 3.52)**	NA	NA	NA	1.25 (0.99, 1.58)	1.25 (0.99, 1.58)
	NNTH	**6 (4, 13)**	NA	NA	NA	NNTH 19 (NNTH 9 to ∞ to NNTB 398)	NNTH 19 (NNTH 9 to ∞ to NNTB 398)
CV death	OR	2.32 (0.78, 8.32)	NA	NA	NA	NA	NA
	NNTH	NNTH 6 (NNTH 3 to ∞ to NNTB 23)	NA	NA	NA	NA	NA
All-cause death	OR	1.89 (0.82, 4.73)	NA	NA	NA	0.88 (0.75, 1.05)	0.88 (0.75, 1.05)
	NNTH	NNTH 17 (NNTH 10 to ∞ to NNTB 39)	NA	NA	NA	NNTB 62 (NNTB 26 to ∞ to NNTH 174)	NNTB 62 (NNTB 26 to ∞ to NNTH 174)
Studies controlled with active therapy							
MI	OR	1.00 (0.36, 2.82)	NA	NA	1.13 (0.98, 1.31)	**1.22 (1.04, 1.44)**	**1.18 (1.06, 1.32)**
	NNTH	NA (NNTH 11 to ∞ to NNTB 5)	NA	NA	NNTH 64 (NNTH 31 to ∞ to NNTB 365)	**41 (24, 193)**	**48 (30, 131)**
Stroke	OR	1.54 (0.29, 10.02)	NA	NA	1.03 (0.82, 1.30)	1.08 (0.81, 1.46)	1.05 (0.88, 1.27)
	NNTH	NNTH 10 (NNTH 3 to ∞ to NNTB 4)	NA	NA	NNTH 151 (NNTH 17 to ∞ to NNTB 23)	NNTH 58 (NNTH 12 to ∞ to NNTB 21)	NNTH 91 (NNTH 19 to ∞ to NNTB 35)
CHF	OR	1.23 (0.47, 3.32)	NA	NA	**1.74 (1.37, 2.20)**	**1.32 (1.11, 1.56)**	**1.43 (1.23, 1.65)**
	NNTH	NNTH 20 (NNTH 4 to ∞ to NNTB 5)	NA	NA	**8 (6, 13)**	**15 (10, 39)**	**12 (9, 20)**
CV death	OR	0.40 (0.04, 2.45)	NA	NA	0.88 (0.69, 1.12)	1.34 (0.98, 1.84)	1.15 (0.82, 1.62)
	NNTH	NNTB 7 (NNTB 4 to ∞ to NNTH 6)	NA	NA	NNTB 44 (NNTB 16 to ∞ to NNTH 48)	NNTH 18 (NNTH 8 to ∞ to NNTB 272)	NNTH 38 (NNTH 11 to ∞ to NNTB 29)
All-cause death	OR	0.79 (0.25, 2.38)	NA	NA	1.16 (0.86, 1.56)	1.06 (0.88, 1.30)	1.09 (0.94, 1.27)
	NNTH	NNTB 32 (NNTB 3 to ∞ to NNTH 14)	NA	NA	NNTH 59 (NNTH 22 to ∞ to NNTB 52)	NNTH 146 (NNTH 35 to ∞ to NNTB 62)	NNTH 100 (NNTH 131 to ∞ to NNTB 131)
Monotherapy studies							
MI	OR	1.36 (0.53, 3.80)	NA	NA	1.56 (0.99, 2.44)	**1.26 (1.01, 1.57)**	**1.31 (1.08, 1.60)**
	NNTH	NNTH 27 (NNTH 9 to ∞ to NNTB 9)	NA	NA	NNTH 20 (NNTH 12 to ∞ to NNTB 737)	**35 (20, 751)**	**31 (19, 100)**
Stroke	OR	1.17 (0.28, 5.69)	NA	NA	1.14 (0.98, 1.33)	1.31 (0.94, 1.83)	**1.17 (1.02, 1.35)**
	NNTH	NNTH 28	NA	NA	NNTH 34	NNTH 16	**28 (15, 225)**

Studies with rosiglitazone		FDA Meta-analysis of RCTs	RECORD Trial - original	RECORD Trial – re-adjudication	Meta-analysis of observational studies		
					Case-control	Cohort	Overall
		(NNTH 3 to ∞ to NNTB 4)			(NNTH 16 to ∞ to NNTB 221)	(NNTH 7 to ∞ to NNTB 72)	
CHF	OR	1.25 (0.43, 3.89)	NA	NA	**1.96 (1.41, 2.72)**	**1.25 (1.02, 1.54)**	1.54 (0.99, 2.37)
	NNTH	NNTH 19 (NNTH 4 to ∞ to NNTB 4)	NA	NA	**7 (5, 12)**	**19 (10, 203)**	NNTH 10 (NNTH 5 to ∞ to NNTB 398)
CV death	OR	0.55 (0.08, 3.44)	NA	NA	0.88 (0.59, 1.31)	1.34 (0.98, 1.84)	1.11 (0.73, 1.67)
	NNTH	NNTB 10 (NNTB 4 to ∞ to NNTH 4)	NA	NA	NNTB 44 (NNTB 12 to ∞ to NNTH 20)	NNTH 18 (NNTH 8 to ∞ to NNTB 272)	NNTH 52 (NNTH 10 to ∞ to NNTB 18)
All-cause death	OR	1.02 (0.33, 3.33)	NA	NA	1.11 (0.71, 1.74)	1.13 (0.87, 1.46)	1.12 (0.90, 1.41)
	NNTH	NNTH 423 (NNTH 4 to ∞ to NNTB 5)	NA	NA	NNTH 83 (NNTH 19 to ∞ to NNTB 21)	NNTH 72 (NNTH 26 to ∞ to NNTB 56)	NNTH 77 (NNTH 28 to ∞ to NNTB 75)
Add-on studies							
MI	OR	2.82 (0.49, 29.32)	1.14 (0.8, 1.63)	1.13 (0.80, 1.59)	1.06 (0.97, 1.15)	1.09 (0.91, 1.32)	1.06 (0.97, 1.16)
	NNTH	NNTH 11 (NNTH 7 to ∞ to NNTB 9)	NNTH 60 (NNTH 19 to ∞ to NNTB 30)	NNTH 64 (NNTH 19 to ∞ to NNTB 30)	NNTH 131 (NNTH 56 to ∞ to NNTB 242)	NNTH 89 (NNTH 30 to ∞ to NNTB 76)	NNTH 131 (NNTH 53 to ∞ to NNTB 241)
Stroke	OR	0.34 (0.01, 4.32)	0.72 (0.49, 1.06)	0.79 (0.54, 1.14)	0.81 (0.59, 1.12)	0.95 (0.60, 1.50)	0.91 (0.64, 1.31)
	NNTH	NNTB 5 (NNTB 2 to ∞ to NNTH 3)	NNTB 14 (NNTB 7 to ∞ to NNTH 76)	NNTB 19 (NNTB 8 to ∞ to NNTH 34)	NNTB 21 (NNTB 9 to ∞ to NNTH 39)	NNTB 87 (NNTB 9 to ∞ to NNTH 11)	NNTB 47 (NNTB 10 to ∞ to NNTH 16)
CHF	OR	1.92 (0.87, 4.39)	**2.10 (1.35, 3.27)**	NA	**1.43 (1.25, 1.63)**	**1.36 (1.18, 1.56)**	**1.39 (1.27, 1.53)**
	NNTH	NNTH 7 (NNTH 4 to ∞ to NNTB 28	**6 (4, 14)**	NA	**12 (9, 18)**	**14 (10, 25)**	**13 (10, 17)**
CV death	OR	2.09 (0.11, 124.53)	0.84 (0.59, 1.18)	0.90 (0.68, 1.21)	NA	NA	NA
	NNTH	NNTH 7 (NNTH 2 to ∞ to NNTB 4)	NNTB 32 (NNTB 12 to ∞ to NNTH 32)	NNTB 53 (NNTB 15 to ∞ to NNTH 28)	NA	NA	NA
All-cause death	OR	1.57 (0.18, 19.10)	0.86 (0.68, 1.08)	0.86 (0.68, 1.08)	**1.26 (1.12, 1.42)**	**0.89 (0.80, 0.98)†**	0.96 (0.80, 1.16)
	NNTH	NNTH 22 (NNTH 8 to ∞ to NNTB 2)	NNTB 52 (NNTB 18 to ∞ to NNTH 111)	NNTB 52 (NNTB 18 to ∞ to NNTH 111)	**40 (27, 77)**	**NNTB 68 (NNTB 35 to NNTB 408)†**	NNTB 200 (NNTB 34 to ∞ to NNTH 60)
Overall							
MI	OR	**1.80 (1.03, 3.25)**	NA	NA	1.13 (0.99, 1.29)	**1.12 (1.04, 1.42)**	**1.17 (1.06, 1.30)**
	NNTH	**16 (10, 255)**	NA	NA	NNTH 64 (NNTH 32 to ∞ to NNTB 738)	**69 (24, 193)**	**51 (32, 131)**
Stroke	OR	0.86 (0.40, 1.83)	NA	NA	1.03 (0.82, 1.30)	1.17 (0.84, 1.62)	1.10 (0.90, 1.33)
	NNTH	NNTB 30 (NNTB 5 to ∞ to NNTH 7)	NA	NA	NNTH 151 (NNTH 17 to ∞ to NNTB 23)	NNTH 28 (NNTH 9 to ∞ to NNTB 26)	NNTH 47 (NNTH 16 to ∞ to NNTB 43)
CHF	OR	**1.93 (1.30, 2.93)**	NA	NA	**1.74 (1.37, 2.20)**	**1.31 (1.14, 1.51)**	**1.41 (1.23 ,1.61)**
	NNTH	**7 (5, 16)**	NA	NA	**8 (6, 13)**	**15 (10, 31)**	**12 (9, 20)**
CV death	OR	1.46 (0.60, 3.77)	NA	NA	0.88 (0.69, 1.12)	1.34 (0.98, 1.84)	1.15 (0.82, 1.62)
	NNTH	NNTH 14 (NNTH 4 to ∞ to NNTB 12)	NA	NA	NNTB 44 (NNTB 16 to ∞ to NNTH 48)	NNTH 18 (NNTH 8 to ∞ to NNTB 272)	NNTH 38 (NNTH 11 to ∞ to NNTB 29)
All-cause	OR	1.38 (0.72, 2.72)	NA	NA	1.16 (0.86, 1.56)	1.04 (0.87, 1.25)	1.07 (0.93, 1.24)

Studies with rosiglitazone		FDA Meta-analysis of RCTs	RECORD Trial - original	RECORD Trial – re-adjudication	Meta-analysis of observational studies		
					Case-control	Cohort	Overall
death	NNTH	NNTH 29 (NNTH 5 to ∞ to NNTB 22)	NA	NA	NNTH 60 (NNTH 22 to ∞ to NNTB 52)	NNTH 215 (NNTH 41 to ∞ to NNTB 56)	NNTH 126 (NNTH 42 to ∞ to NNTB 111)

CHF, congestive heart failure; CV, cardiovascular; FDA, Food and Drug Administration; MI, myocardial infarction; NA, not applicable; NNH, number needed to harm; OR, odds ratio; RCT, randomized controlled trial.
Bold values are statistically significant (95% Confidence Interval).
† Statistically significant values of NNTB indicate a protective effect.
The baseline event rates per year for CV events used in the calculation of NNH values were obtained from the Look AHEAD Research Group Trial (all-cause death, 0.86%; CV death, 0.24%; MI, 0.84%; stroke, 0.34%; and CHF, 0.51%). The mean follow-up on thiazolidinedione was of 188 days in the FDA Meta-analysis of RCTs, 5.5 years in the RECORD Trial, and ranged between 105 days and 7.1 years in studies included in the Meta-analysis of observational studies.

Table II. 2 – Adjusted odds ratio (95% confidence intervals) and number needed to treat to harm (95% confidence intervals) for cardiovascular adverse events associated with the use of pioglitazone in several settings according to each type of evidence.

Studies with pioglitazone		FDA Meta-analysis of RCTs	PROACTIVE	Meta-analysis of observational studies		
				Case-control	Cohort	Overall
Studies controlled with placebo						
MI	OR	0.41 (0.09, 1.56)	NA	1.21 (0.87, 1.67)	0.71 (0.39, 1.30)	0.99 (0.59, 1.64)
	NNTH	NNTB 6 (NNTB 1 to ∞ to NNTH 20)	NA	NNTH 42 (NNTH 18 to ∞ to NNTB 50)	NNTB 19 (NNTB 5 to ∞ to NNTH 32)	NNTB 737 (NNTB 11 to ∞ to NNTH 18)
Stroke	OR	1.64 (0.08, 99.71)	NA	NA	NA	NA
	NNTH	NNTH 9 (NNTH 2 to ∞ to NNTB 3)	NA	NA	NA	NA
CHF	OR	1.77 (0.62, 5.75)	NA	NA	1.25 (0.99, 1.58)	1.25 (0.99, 1.58)
	NNTH	NNTH 8 (NNTH 3 to ∞ to NNTB 8)	NA	NA	NNTH 19 (NNTH 9 to ∞ to NNTB 398)	NNTH 19 (NNTH 9 to ∞ to NNTB 398)
CV death	OR	0.80 (0.10, 6.14)	NA	NA	NA	NA
	NNTH	NNTB 26 (NNTB 4 to ∞ to NNTH 3)	NA	NA	NA	NA
All-cause death	OR	0.63 (0.12, 3.01)	NA	NA	**0.69 (0.49, 0.98)**	**0.69 (0.49, 0.98)**
	NNTH	NNTB 15 (NNTB 2 to ∞ to NNTH 12)	NA	NA	**NNTB 19 (NNTB 9 to NNTB 408)†**	**NNTB 19 (NNTB 9 to NNTB 408)†**
Studies controlled with active therapy						
MI	OR	1.08 (0.60, 1.94)	NA	0.96 (0.80, 1.15)	0.80 (0.60, 1.05)	0.85 (0.69, 1.04)
	NNTH	NNTH 100 (NNTH 15 to ∞ to NNTB 12)	NA	NNTB 179 (NNTH 56 to ∞ to NNTB 30)	NNTB 30 (NNTB 12 to ∞ to NNTH 156)	NNTB 43 (NNTB 17 to ∞ to NNTH 193)
Stroke	OR	0.53 (0.19, 1.34)	NA	0.89 (0.49, 1.60)	0.88 (0.75, 1.04)	0.88 (0.75, 1.03)
	NNTH	NNTB 7 (NNTB 3 to ∞ to NNTH 15)	NA	NNTB 39 (NNTB 7 to ∞ to NNTH 9)	NNTB 35 (NNTB 16 to ∞ to NNTH 113)	NNTB 35 (NNTB 16 to ∞ to NNTH 151)
CHF	OR	1.44 (0.96, 2.19)	NA	1.07 (0.89, 1.28)	0.92 (0.66, 1.29)	0.93 (0.70, 1.23)
	NNTH	NNTH 12 (NNTH 6 to ∞ to NNTB 97)	NA	NNTH 60 (NNTH 17 to ∞ to NNTB 34)	NNTB 47 (NNTB 9 to ∞ to NNTH 16)	NNTB 55 (NNTB 11 to ∞ to NNTH 20)
CV death	OR	1.26 (0.60, 2.67)	NA	NA	1.37 (0.77, 2.43)	1.37 (0.77, 2.43)
	NNTH	NNTH 23 (NNTH 5 to ∞ to NNTB 12)	NA	NA	NNTH 17 (NNTH 6 to ∞ to NNTB 22)	NNTH 17 (NNTH 6 to ∞ to NNTB 22)
All-cause death	OR	1.17 (0.64, 2.14)	NA	1.15 (0.96, 1.38)	0.69 (0.44, 1.10)	0.77 (0.49, 1.20)
	NNTH	NNTH 56 (NNTH 15 to ∞ to NNTB 15)	NA	NNTH 63 (NNTH 29 to ∞ to NNTB 200)	NNTB 19 (NNTB 7 to ∞ to NNTH 91)	NNTB 28 (NNTB 9 to ∞ to NNTH 49)
Monotherapy studies						
MI	OR	0.71 (0.29, 1.67)	NA	0.73 (0.50, 1.35)	**0.84 (0.71, 0.99)**	**0.83 (0.71, 0.98)**
	NNTH	NNTB 19 (NNTB 4 to ∞ to NNTH 18)	NA	NNTB 21 (NNTB 6 to ∞ to NNTH 28)	**NNTB 40 (NNTB 19 to NNTB 737)†**	**NNTB 37 (NNTB 19 to NNTB 365)†**
Stroke	OR	1.50 (0.35, 7.22)	NA	1.25 (0.61, 2.55)	0.92 (0.72, 1.19)	0.95 (0.75, 1.21)
	NNTH	NNTH 11 (NNTH 3 to ∞ to NNTB 5)	NA	NNTH 20 (NNTH 5 to ∞ to NNTB 9)	NNTB 54 (NNTB 14 to ∞ to NNTH 25)	NNTB 87 (NNTB 16 to ∞ to NNTH 23)
CHF	OR	1.20 (0.65, 2.24)	NA	0.91 (0.52, 1.59)	1.10 (0.86, 1.39)	1.07 (0.86, 1.35)
	NNTH	NNTH 23 (NNTH 6 to ∞ to NNTB 9)	NA	NNTB 43 (NNTB 6 to ∞ to NNTH 9)	NNTH 43 (NNTH 13 to ∞ to NNTB 26)	NNTH 60 (NNTH 14 to ∞ to NNTB 26)
CV death	OR	1.33 (0.48, 3.86)	NA	NA	1.37 (0.77, 2.43)	1.37 (0.77, 2.43)
	NNTH	NNTH 18 (NNTH 4 to ∞ to NNTB 9)	NA	NA	NNTH 17 (NNTH 6 to ∞ to NNTB 22)	NNTH 17 (NNTH 6 to ∞ to NNTB 22)
All-cause death	OR	0.83 (0.37, 1.83)	NA	0.94 (0.44, 2.00)	0.82 (0.57, 1.16)	0.84 (0.61, 1.15)
	NNTH	NNTB 41 (NNTB 9 to ∞ to NNTH 6)	NA	NNTB 131 (NNTB 7 to ∞ to NNTH 16)	NNTB 39 (NNTB 12 to ∞ to NNTH 60)	NNTB 44 (NNTB 14 to ∞ to NNTH 63)
Add-on studies						

Studies with pioglitazone		FDA Meta-analysis of RCTs	PROACTIVE	Meta-analysis of observational studies		
				Case-control	Cohort	Overall
MI	OR	0.57 (0.12, 2.25)	0.83 (0.65, 1.06)	1.02 (0.87, 1.19)	0.76 (0.50, 1.14)	0.87 (0.68, 1.12)
	NNTH	NNTB 11 (NNTB 2 to ∞ to NNTH 13)	NNTB 37 (NNTB 15 to ∞ to NNTH 131)	NNTH 379 (NNTH 46 to ∞ to NNTB 50)	NNTB 24 (NNTB 8 to ∞ to NNTH 60)	NNTB 50 (NNTB 17 to ∞ to NNTH 69)
Stroke	OR	0.41 (0.00 , 5.33)	0.81 (0.61, 1.07)	0.68 (0.38, 1.20)	0.80 (0.59, 1.09)	0.80 (0.63, 1.02)
	NNTH	NNTB 6 (NA to ∞ to NNTH 3)	NNTB 21 (NNTB 9 to ∞ to NNTH 66)	NNTB 12 (NNTB 5 to ∞ to NNTH 24)	NNTB 20 (NNTB 9 to ∞ to NNTH 52)	NNTB 20 (NNTB 10 to ∞ to NNTH 225)
CHF	OR	1.40 (0.64, 3.15)	**1.41 (1.10, 1.80)**	1.09 (0.90, 1.32)	0.84 (0.52, 1.36)	0.87 (0.58, 1.32)
	NNTH	NNTH 13 (NNTH 4 to ∞ to NNTB 9)	**12 (8, 43)**	NNTH 47 (NNTH 15 to ∞ to NNTB 37)	NNTB 22 (NNTB 6 to ∞ to NNTH 14)	NNTB 28 (NNTB 7 to ∞ to NNTH 15)
CV death	OR	1.51 (0.17, 18.18)	0.94 (0.74, 1.20)	NA	NA	NA
	NNTH	NNTH 13 (NNTH 2 to ∞ to NNTB 5)	NNTB 90 (NNTB 19 to ∞ to NNTH 29)	NA	NA	NA
All-cause death	OR	1.34 (0.23, 9.21)	0.96 (0.78, 1.18)	1.20 (0.98, 1.47)	0.59 (0.34, 1.03)	0.69 (0.37, 1.29)
	NNTH	NNTH 32 (NNTH 9 to ∞ to NNTB 3)	NNTB 200 (NNTB 30 to ∞ to NNTH 54)	NNTH 49 (NNTH 25 to ∞ to NNTB 408)	NNTB 13 (NNTB 5 to ∞ to NNTH 284)	NNTB 19 (NNTB 6 to ∞ to NNTH 36)
Overall						
MI	OR	0.91 (0.53, 1.53)	NA	0.99 (0.86, 1.16)	0.79 (0.59, 1.05)	0.87 (0.71, 1.06)
	NNTH	NNTB 76 (NNTB 9 to ∞ to NNTH 21)	NA	NNTB 737 (NNTB 46 to ∞ to NNTH 53)	NNTB 29 (NNTB 11 to ∞ to NNTH 156)	NNTB 50 (NNTB 19 to ∞ to NNTH 131)
Stroke	OR	0.61 (0.24, 1.43)	NA	0.89 (0.49, 1.60)	0.92 (0.77, 1.09)	0.91 (0.76, 1.07)
	NNTH	NNTB 9 (NNTB 4 to ∞ to NNTH 12)	NA	NNTB 39 (NNTB 7 to ∞ to NNTH 9)	NNTB 54 (NNTB 17 to ∞ to NNTH 52)	NNTB 47 (NNTB 17 to ∞ to NNTH 66)
CHF	OR	**1.47 (1.01, 2.16)**	NA	1.07 (0.89, 1.28)	0.94 (0.68, 1.28)	0.94 (0.72, 1.23)
	NNTH	**11 (6, 403)**	NA	NNTH 60 (NNTH 17 to ∞ to NNTB 34)	NNTB 64 (NNTB 10 to ∞ to NNTH 17)	NNTB 64 (NNTB 12 to ∞ to NNTH 20)
CV death	OR	1.18 (0.60, 2.34)	NA	NA	1.37 (0.77, 2.43)	1.37 (0.77, 2.43)
	NNTH	NNTH 32 (NNTH 5 to ∞ to NNTB 12)	NA	NA	NNTH 17 (NNTH 6 to ∞ to NNTB 22)	NNTH 17 (NNTH 6 to ∞ to NNTB 22)
All-cause death	OR	1.06 (0.61, 1.85)	NA	1.15 (0.96, 1.38)	0.69 (0.44, 1.10)	0.76 (0.48, 1.20)
	NNTH	NNTH 146 (NNTH 17 to ∞ to NNTB 14)	NA	NNTH 63 (NNTH 30 to ∞ to NNTB 200)	NNTB19 (NNTB 7 to ∞ to NNTH 91)	NNTB 27 (NNTB 8 to ∞ to NNTH 49)

CHF, congestive heart failure; CV, cardiovascular; FDA, Food and Drug Administration; MI, myocardial infarction; NA, not applicable; NNTB, number needed to treat to benefit; NNTH, number needed to treat to harm; OR, odds ratio; RCT, randomized controlled trial.
Bold values are statistically significant (95% Confidence Interval).
† Statistically significant values of NNTB indicate a protective effect.
The baseline event rates per year for CV events used in the calculation of NNH values were obtained from the Look AHEAD Research Group Trial (all-cause death, 0.86%; CV death, 0.24%; MI, 0.84%; stroke, 0.34%; and CHF, 0.51%). The mean follow-up on thiazolidinedione was of 265 days in the FDA Meta-analysis of RCTs, 34.5 months in the PROACTIVE Trial, and ranged between 105 days and 7.1 years in studies included in the Meta-analysis of observational studies.

Table II. 3 – Adjusted odds ratio (95% confidence intervals) and number needed to treat to harm (95% confidence intervals) for cardiovascular adverse events associated with the use of rosiglitazone versus pioglitazone in several settings.

Rosiglitazone vs. pioglitazone		Meta-analysis of observational studies		
		Case-control	Cohort	Overall
Monotherapy studies				
MI	OR	NA	1.23 (0.75, 2.01)	1.23 (0.75, 2.01)
	NNTH	NA	NNTH 39 (NNTH to ∞ to NNTB 23)	NNTH 39 (NNTH 14 to ∞ to NNTB 23)
Stroke	OR	NA	1.33 (0.89, 1.98)	1.33 (0.89, 1.98)
	NNTH	NA	NNTH 16 (NNTH 7 to ∞ to NNTB 39)	NNTH 16 (NNTH 7 to ∞ to NNTB 39)
CHF	OR	NA	0.84 (0.52, 1.35)	0.84 (0.52, 1.35)
	NNTH	NA	NNTB 22 (NNTB 6 to ∞ to NNTB 14)	NNTB 22 (NNTB 6 to ∞ to NNTH 14)
CV death	OR	NA	0.93 (0.21, 4.12)	0.93 (0.21, 4.12)
	NNTH	NA	NNTB 77 (NNTB 5 to ∞ to NNTH 4)	NNTB 77 (NNTB 5 to ∞ to NNTH 4)
All-cause death	OR	NA	1.06 (0.64, 1.74)	1.06 (0.64, 1.74)
	NNTH	NA	NNTH 146 (NNTH 19 to ∞ to NNTB 16)	NNTH 146 (NNTH 19 to ∞ to NNTB 16)
Add-on studies				
MI	OR	1.13 (0.77, 1.65)	**1.11 (1.00, 1.22)**	**1.11 (1.01, 1.22)**
	NNTH	NNTH 64 (NNTH 18 to ∞ to NNTB 26)	**74 (41, 2487)**	**74 (41, 751)**
Stroke	OR	NA	**1.12 (1.00, 1.24)**	**1.12 (1.00, 1.24)**
	NNTH	NA	**39 (21, 4458)**	**39 (21, 4458)**
CHF	OR	NA	**1.15 (1.04, 1.26)**	**1.15 (1.04, 1.26)**
	NNTH	NA	**29 (18, 103)**	**29 (18, 103)**
CV death	OR	NA	NA	NA
	NNTH	NA	NA	NA
All-cause death	OR	NA	**1.15 (1.09, 1.20)**	**1.15 (1.09, 1.20)**
	NNTH	NA	**63 (49, 100)**	**63 (49, 100)**
Overall				
MI	OR	1.12 (0.78, 1.59)	**1.12 (1.01, 1.25)**	**1.12 (1.02, 1.30)**
	NNTH	NNTH 69 (NNTH 19 to ∞ to NNTB 27)	**69 (37, 751)**	**69 (32, 379)**
Stroke	OR	NA	**1.13 (1.02, 1.25)**	**1.13 (1.02, 1.25)**
	NNTH	NA	**36 (20, 225)**	**36 (20, 225)**
CHF	OR	NA	**1.13 (1.09, 1.25)**	**1.13 (1.09, 1.25)**
	NNTH	NA	**33 (19, 47)**	**33 (19, 47)**
CV death	OR	NA	0.93 (0.21, 4.12)	0.93 (0.21, 4.12)
	NNTH	NA	NNTB 77 (NNTB 5 to ∞ to NNTH 4)	NNTB 77 (NNTB 5 to ∞ to NNTH 4)
All-cause death	OR	NA	**1.15 (1.09, 1.20)**	**1.15 (1.09, 1.20)**
	NNTH	NA	**63 (49, 100)**	**63 (49, 100)**

CHF, congestive heart failure; CV, cardiovascular; MI, myocardial infarction; NA, not applicable; NNTB, number needed to treat to benefit; NNTH, number needed to treat to harm; OR, odds ratio.
Bold values are statistically significant (95% Confidence Interval).
The baseline event rates per year for CV events used in the calculation of NNTH values were obtained from the Look AHEAD Research Group Trial (all-cause death, 0.86%; CV death, 0.24%; MI, 0.84%; stroke, 0.34%; and CHF, 0.51%) (Look AHEAD Research Group 2013). The mean follow-up on thiazolidinediones ranged between 105 days and 7.1 years in studies included in the Meta-analysis of observational studies.

II.4.2. CARDIOVASCULAR DEATH

Neither rosiglitazone nor pioglitazone increased the risk of CV death. No differences were found between rosiglitazone and pioglitazone.

II.4.3. MYOCARDIAL INFARCTION

Results from meta-analyses of RCTs indicated an increased risk of MI with rosiglitazone when all studies were considered (NNTH 16, 95% CI 10–255) and also when only placebo-controlled studies were included (NNTH 13, 95% CI 9–60). Meta-analyses of observational studies also found an increased risk of MI with rosiglitazone versus non-TZD comparators in several settings (overall, NNTH 51, 95% CI 32–131; versus only active comparators, NNTH 48, 95% CI 30–131; and only monotherapy studies, NNTH 31, 95% CI 19–100). Rosiglitazone also increased the risk of MI when directly compared with pioglitazone in meta-analyses of observational studies (overall, NNTH 69, 95% CI 32–379; and only add-on studies, NNTH 74, 95% CI 41–751).

II.4.4. STROKE

Only meta-analyses of observational studies revealed an increased risk of stroke with rosiglitazone versus non-TZD comparators in monotherapy studies (NNTH 28, 95% CI 15–225). Compared with pioglitazone, rosiglitazone was associated with an increased risk of stroke (overall, NNTH 36, 95% CI 20–225; and only add-on studies, NNTH 39, 95% CI 21–4458).

II.4.5. CONGESTIVE HEART FAILURE

Meta-analyses of RCTs revealed an increased risk of CHF with rosiglitazone versus non-TZD comparators (NNTH 7, 95% CI 5–16) and versus only placebo (NNTH 6, 95% CI 4–13). Pioglitazone was also associated with an increased risk of CHF versus non-TZD comparators (NNTH 11, 95% CI 6–403). The RECORD clinical trial indicated an increased risk of CHF with rosiglitazone added to metformin or sulfonylurea versus metformin in combination with sulfonylurea (NNTH 6, 95% CI 4–14). The PROACTIVE clinical trial

revealed an increased risk of CHF with pioglitazone versus placebo (NNTH 12, 95% CI 8–43) in patients receiving background antidiabetic therapies. Meta-analyses of observational studies revealed an increased risk of CHF with rosiglitazone versus non-TZD comparators (overall, NNTH 12, 95% CI 9–20; versus only active comparators, NNTH 12, 95% CI 9–20; only add-on studies, NNTH 13, 95% CI 10–17). Compared with pioglitazone, rosiglitazone increased the risk of CHF according to the overall results (NNTH 33, 95% CI 19–47) and when only add-on studies were considered (NNTH 29, 95% CI 18–103).

II.5. DISCUSSION

Rosiglitazone and its CV safety have been discussed in medical literature since the publication of a meta-analysis of RCTs, by Nissen and Wolski, indicating a statistically significant increased risk of MI and a trend toward increased mortality (Nissen & Wolski 2007), (Psaty & Furberg 2007), (Krall 2007), (Drazen, Morrissey & Curfman 2007), (Cleland & Atkin 2007), ([No authors listed] 2007), (Bloomgarden 2007), (Shuster & Schatz 2008). Concerns about the safety of rosiglitazone prompted the re-evaluation of its benefit–risk ratio by regulatory authorities. The FDA and the EMA analysed the data from meta-analyses of short-term RCTs, isolated long-term RCTs, including the RECORD trial, and observational studies designed to assess the risk of CV adverse events in patients taking rosiglitazone or pioglitazone (FDA 2010a), (FDA 2013b), (EMA 2010c). Although both agencies had analysed the same information, their decisions were not coincident. In 2010, rosiglitazone was withdrawn from the market in Europe while the USA imposed restrictions to its utilization (EMA 2010b), (FDA 2011). Those restrictions were eased after the analysis of the results obtained in the readjudication of CV adverse events within the RECORD trial (FDA 2013c).

Although risk assessments may involve quantitative analyses, its key component is a subjective qualitative weighing of the evidence relying on expert opinions (FDA 2013a), (Curtin & Schulz 2011). The introduction of metric indices into this process may contribute to improve the objectivity and reproducibility of regulatory decisions on drug safety, in the light of the rosiglitazone case.

There are a variety of measures of effect size that can be used to describe differences between interventions. Relative measures of potential benefit or potential harm, such as the RR, OR, and HR, are commonly seen in the medical literature (Citrome 2010). However, relative measures do not reflect the risk of the outcome of interest without therapy (baseline risk), and therefore it is not possible to discriminate huge treatment effects from small one (Straus et al. 2011). For example, if the rate of a given adverse event is trivial (0.003%) or meaningful (30%) in the experimental group and similarly trivial (0.001%) or meaningful (10%) in the control group, the RR will always be 3.0. Although the difference is statistically significant for both scenarios, the clinical relevance varies. Decision makers need to know how often this difference in risk is encountered in day-to-day clinical practice (Citrome 2010). In order to answer this question, absolute measures, such as NNTH, are needed. Using the example provided earlier, NNTH would range between 50000 and 5

depending on the scenario in analysis. This example illustrates the potential usefulness of metric indices for making decisions on drug utilization.

The findings of present study point out an increased risk of MI and CHF associated with the use of rosiglitazone versus comparators in both meta-analysis of short-term RCTs and meta-analysis of observational studies. The overall results of the meta-analysis of short-term RCTs estimated NNTH values at 16 for MI and 7 for CHF. According to the overall results of the meta-analysis of observational studies, NNTH values were found to be 51 for MI and 12 for CHF. The results from the RECORD trial indicated an increased risk for CHF (NNTH 6) but not for MI. Furthermore, when rosiglitazone was directly compared with pioglitazone in the meta-analysis of observational studies, a statistically significant increased risk of MI, stroke, CHF, and all-cause death was found in patients treated with rosiglitazone, with NNTH values lower than 70 irrespectively of the CV adverse event. The results obtained across several sources of evidence are consistent with an increased CV risk in patients receiving rosiglitazone compared with other antidiabetics, including pioglitazone.

There are several issues that must be taken into account when one is interpreting the results found in here. First, NNTH values were estimated by applying ORs from meta-analyses to baseline event rates per year. The baseline event rates were those of overweight or obese patients with type 2 diabetes allocated to the control group of a single RCT (Look AHEAD Research Group 2013). The average age of patients was 59 years, the median duration of diabetes was 5 years, and 14% of patients reported a history of CV disease (Look AHEAD Research Group 2013). Because NNTH estimates vary according to baseline event rates, the results of the present study are applicable only to populations with similar characteristics (Laupacis, Sackett & Roberts 1988), (Baglin 2009). This is a limitation of the methodology that precludes the generalization of the results to all patients. Second, studies included in the several meta-analyses have intrinsic limitations that are insurmountable. Because RCTs included in the FDA meta-analysis were not designed to assess CV adverse events, an incorrect adjudication of events can lead to misleading estimates of risk (FDA 2010a), (FDA 2013b), (EMA 2010c). Patients that received rosiglitazone were possibly at a more advanced stage of diabetes with a higher risk for harmful events compared with those on pioglitazone because of a longer duration of disease (7 versus 6 years) (FDA 2010a), (FDA 2013b), (EMA 2010c). Further, summary estimates of NNTH assume constant risk differences between studies, a challenging assumption because of inevitable variation in baseline event rates between studies, and differences in duration of follow-up (i.e., time horizon) (Marx & Bucher 2003). It must be noted that the duration of follow-up varied

between 105 days and 7 years in studies included in the meta-analyses of observational studies. RRs for adverse events may vary with different durations of follow-up. As an example, Nissen and Wolski noted that excluding the long-term RECORD trial from their meta-analysis resulted in a numerically higher OR, although they have reached similar conclusions (Nissen & Wolski 2010). Nevertheless, evidence suggests that RR and OR provide more homogenous estimates than absolute risk differences (McAlister 2002). Third, the RECORD trial, which was designed to assess CV outcomes after the addition of rosiglitazone to either metformin or sulfonylurea, also has limitations. The results confirmed an increased risk of CHF with rosiglitazone but did not confirm nor ruled out a possible increased risk of MI (Home et al. 2009). However, the RECORD trial had an open-label design and a smaller sample size than other trials designed to assess CV outcomes (Bourg & Phillips 2012). Additionally, it was noted a lower event rate than the expected and high annual loss to follow-up, which decreases the statistical power of the trial (Bourg & Phillips 2012).

Post-marketing drug risk assessment should integrate evidence resulting from several sources of data. For that reason, several meta-analyses, including both interventional and observational data, were considered in this study for estimating NNTH values in different scenarios. However, the inclusion of observational studies in meta-analysis may lead to an increase in the between-studies heterogeneity, as it was observed in this study (Alves, Batel-Marques & Macedo 2014), (Berlin, Soledad-Cepeda & Kim 2012). Nevertheless, evidence from observational studies should not be dismissed (Vandenbroucke 2004), (Vandenbroucke 2006). The extent to which different study designs contribute to the benefit–risk ratio evaluation of drugs deserves further considerations. Experts from the EMA have recognized the additional value of observational data over RCTs in supporting post-marketing drug safety evaluations (EMA 2010c). Observational studies are more likely to detect rare and long-term latency adverse events, and this type of data may represent better the frequency of harmful effects experienced in actual clinical practice (Vandenbroucke & Psaty 2008), (AHRQ 2014), (Vandenbroucke 2004), (Vandenbroucke 2006), (Glasziou, Vandenbroucke & Chalmers 2004), (Papanikolaou, Christidi & Ioannidis 2006). A previous study using the CV toxicity associated with selective COX-2 inhibitors, for example, concluded that rigid classification of evidence is not appropriate in monitoring risks and benefits and that all valid evidence needs to be included, beyond RCTs (van Staa et al. 2008). The value given by each regulatory authority to different types of study designs may help to explain the different decisions made by the EMA and the FDA.

There are limitations in the context of benefit–risk assessment that are not overcome by the application of a quantitative methodology, such as NNTB and NNTH. The application of quantitative metrics does not intend to replace the qualitative assessment, which relies on scientific and clinical judgment. However, the establishment of structured frameworks for benefit–risk assessment, which comprises qualitative and quantitative approaches, can contribute to improve transparency and traceability of regulatory decisions (Nixon et al. 2016), (Mt-Isa et al. 2014), (Holden, Juhaeri & Dai 2003a), (Holden, Juhaeri & Dai 2003). Quantitative methodologies, in particular, allow that sensitivity analyses can be carried out to assess the impact of different assumptions on the benefit–risk ratio conclusions (Nixon et al. 2016), (Mt-Isa et al. 2014), (Hallgreen et al. 2014). When a drug is being evaluated, some quantitative approach for assessing benefits and risks may be of increased value and help inform regulatory decisions (Nixon et al. 2016), (Yuan, Levitan & Berlin 2011), (Mt-Isa et al. 2014), (Hallgreen et al. 2014). NNTB/NNTH methodology and derived concepts, such as the weighted net clinical benefit, are well known in medical literature, are easy to understand and communicate, and have proven to be valuable in quantifying benefits and risks of drugs (Nixon et al. 2016). Researchers from the PROTECT Consortium concluded that the simplicity of NNTB/NNTH provides an attractive feature for benefit–risk assessment and recommended further investigation on their usefulness (Mt-Isa et al. 2013), (Mt-Isa et al. 2014). The establishment of thresholds of risk based, for example, on metric indices could be used as an aid in the decision-making process.

Owing to the rosiglitazone case, both agencies proposed draft guidance for data requirements concerning the safety profiles of new antidiabetics (EMA 2012), (FDA 2008a), (FDA 2008b). The guidance published by the FDA defines an explicit level of increased risk of MI from new antidiabetics: upper bound of two-sided 95% CI of risk ratios of 1.8 for premarketing and 1.3 for post-marketing trials (FDA 2008a). The EMA did not specify levels of risk and, instead, defined which elements from the drugs' development program would be considered to support the evaluation of the possible excess CV risk. Although both guidelines have been developed with the same purpose, there are differences between them, which may illustrate uncertainties that both authorities faced deciding on rosiglitazone.

The purpose of this study is not to argue in favour or against decisions made by regulatory authorities about rosiglitazone but rather to evaluate the appropriateness of a quantitative approach for benefit–risk assessment. In this particular case, a quantitative approach was not mentioned in the assessment reports produced by both regulatory authorities. According to the findings of this study, NNTH values indicated, in a consistent

way across different sources of evidence, an increased risk of CV adverse events with rosiglitazone versus pioglitazone. Given the severity of the adverse events, the NNTH values may be too low to be acceptable (Citrome & Ketter 2013). However, the establishment of a tolerability of risk threshold based on the NNTH concept would be needed before making a definite conclusion.

This study has demonstrated that NNTH can be used in the context of benefit-risk analysis. However, this quantitative methodology does not replace scientific and clinical judgment, particularly in the light of intrinsic limitations of the studies used to generate risk estimates. The addition of objective and validated metrics, such as NNTB and NNTH, to post-marketing drug's benefit–risk ratio assessment process could be of increased value and help regulatory authorities to make consistent and reproducible decisions on drug safety. Further investigation should be carried out about the role of metric indices in safety assessments in the context of benefit–risk re-evaluations of marketed drugs.

II.6. REFERENCES

[NO AUTHORS LISTED], 2007. ENSURING DRUG SAFETY: LESSONS FROM THE THIAZOLIDINEDIONES. LANCET, 370(9593), P. 1101.

AHRQ, 2014. METHODS GUIDE FOR EFFECTIVENESS AND COMPARATIVE EFFECTIVENESS REVIEWS (PUBLICATION NO. 10(14)-EHC063-EF), ROCKVILLE, MD, USA: AGENCY FOR HEALTHCARE RESEARCH AND QUALITY.

ALTMAN, D., 1998. CONFIDENCE INTERVALS FOR THE NUMBER NEEDED TO TREAT. BMJ, 317(7168), PP. 1309-1312.

ALVES, C., BATEL-MARQUES, F. & MACEDO, A., 2012. A META-ANALYSIS OF SERIOUS ADVERSE EVENTS REPORTED WITH EXENATIDE AND LIRAGLUTIDE: ACUTE PANCREATITIS AND CANCER. DIABETES RES CLIN PRACT, 98(2), PP. 271-284.

ALVES, C., BATEL-MARQUES, F. & MACEDO, A., 2014. DRUG-SAFETY ALERTS ISSUED BY REGULATORY AUTHORITIES: USEFULNESS OF META-ANALYSIS IN PREDICTING RISKS EARLIER. EUR J CLIN PHARMACOL, 70(6), PP. 745-756.

BAGLIN, T., 2009. COMMUNICATING BENEFIT AND RISK. BR J HAEMATOL, 146(1), PP. 31-33.

BERLIN, J., SOLEDAD-CEPEDA, M. & KIM, C., 2012. THE USE OF META-ANALYSIS IN PHARMACOEPIDEMIOLOGY. IN: B. L. STROM, S. E. KIMMEL & S. HENNESSY, EDS. PHARMACOEPIDEMIOLOGY. 5TH ED. CHICHESTER, ENGLAND: JOHN WILEY & SONS, PP. 723-756.

BLOOMGARDEN, Z., 2007. THE AVANDIA DEBATE. DIABETES CARE, 30(9), PP. 2401-2408.

BOURG, C. & PHILLIPS, B., 2012. ROSIGLITAZONE, MYOCARDIAL ISCHEMIC RISK, AND RECENT REGULATORY ACTIONS. ANN PHARMACOTHER, 46(2), PP. 282-289.

CITROME, L., 2010. RELATIVE VS. ABSOLUTE MEASURES OF BENEFIT AND RISK: WHAT'S THE DIFFERENCE? ACTA PSYCHIATR SCAND, 121(2), PP. 94-102.

CITROME, L. & KETTER, T., 2013. WHEN DOES A DIFFERENCE MAKE A DIFFERENCE? INTERPRETATION OF NUMBER NEEDED TO TREAT, NUMBER NEEDED TO HARM, AND LIKELIHOOD TO BE HELPED OR HARMED. INT J CLIN PRACT, 67(5), PP. 407-411.

CLARKE, A., DEEKS, J. & SHAKIR, S., 2006. AN ASSESSMENT OF THE PUBLICLY DISSEMINATED EVIDENCE OF SAFETY USED IN DECISIONS TO WITHDRAW MEDICINAL PRODUCTS FROM THE UK AND US MARKETS. DRUG SAF, 29(2), PP. 175-181.

CLELAND, J. & ATKIN, S., 2007. THIAZOLIDINEDIONES, DEADLY SINS, SURROGATES, AND ELEPHANTS. LANCET, 370(9593), PP. 1103-1104.

COOK, R. & SACKETT, D., 1995. THE NUMBER NEEDED TO TREAT: A CLINICALLY USEFUL MEASURE OF TREATMENT EFFECT. BMJ, 310(6977), PP. 452-454.

CURTIN, F. & SCHULZ, P., 2011. ASSESSING THE BENEFIT: RISK RATIO OF A DRUG--RANDOMIZED AND NATURALISTIC EVIDENCE. DIALOGUES CLIN NEUROSCI, 13(2), PP. 183-190.

DERSIMONIAN, R. & LAIRD, N., 1986. META-ANALYSIS IN CLINICAL TRIALS. CONTROL CLIN TRIALS. 1986, 7(3), PP. 177-188.

DRAZEN, J., MORRISSEY, S. & CURFMAN, G., 2007. ROSIGLITAZONE - CONTINUED UNCERTAINTY ABOUT SAFETY. N ENGL J MED, 357(1), PP. 63-64.

EICHLER, H. ET AL., 2013. THE RISKS OF RISK AVERSION IN DRUG REGULATION. NAT REV DRUG DISCOV, 12(12), PP. 907-916.

EMA, 2007. COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE (CHMP). REFLECTION PAPER ON BENEFIT-RISK ASSESSMENT METHODS IN THE CONTEXT OF THE EVALUATION OF MARKETING AUTHORISATION APPLICATIONS OF MEDICINAL PRODUCTS FOR HUMAN USE (EMEA/CHMP/15404/2007). [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REGULATORY_AND_PROCEDURAL_GUIDELINE/2010/01/WC500069634.PDF [ACCESSED 10 MAY 2016].

EMA, 2010B. EUROPEAN MEDICINES AGENCY RECOMMENDS SUSPENSION OF AVANDIA, AVANDAMET AND AVAGLIM - ANTI-DIABETES MEDICATION TO BE TAKEN OFF THE MARKET [PRESS RELEASE: 23-09-2010]. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/EMA/INDEX.JSP?CURL=PAGES/NEWS_AND_EVENTS/NEWS/2010/09/NEWS_DETAIL_001119.JSP&MID=WC0B01AC058004D5C1[ACCESSED 23 APRIL 2015].

EMA, 2010C. ASSESSMENT REPORT FOR AVANDIA (PROCEDURE NO EMA/H/C/000268/A20/0075). [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/EPAR_-

_ASSESSMENT_REPORT_-_VARIATION/HUMAN/000268/WC500100757.PDF [ACCESSED 23 APRIL 2015].

EMA, 2011B. BENEFIT–RISK METHODOLOGY PROJECT. WORK PACKAGE 3 REPORT: FIELD TESTS. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REPORT/2011/09/WC500112088.PDF [ACCESSED 29 DEC 2016].

EMA, 2012. GUIDELINE ON CLINICAL INVESTIGATION OF MEDICINAL PRODUCTS IN THE TREATMENT OF DIABETES MELLITUS - DRAFT. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/SCIENTIFIC_GUIDELINE/2012/06/WC500129256.PDF [ACCESSED 02 APRIL 2015].

FDA, 2008A. GUIDANCE FOR INDUSTRY: DIABETES MELLITUS - EVALUATING CARDIOVASCULAR RISK IN NEW ANTIDIABETIC THERAPIES TO TREAT TYPE 2 DIABETES. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DOWNLOADS/DRUGS/GUIDANCECOMPLIANCEREGULATORYINFORMATION/GUIDANCES/UCM071627.PDF?UTM_CAMPAIGN=GOOGLE2UTM_SOURCE=FDASEARCHUTM_MEDIUM=WEBSITEUTM_TERM=GUIDANCE [ACCESSED 02 APRIL 2015].

FDA, 2008B. GUIDANCE FOR INDUSTRY. DIABETES MELLITUS: DEVELOPING DRUGS AND THERAPEUTIC BIOLOGICS FOR TREATMENT AND PREVENTION - DRAFT. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DOWNLOADS/DRUGS/GUIDANCES/UCM071624.PDF [ACCESSED 02 APRIL 2015].

FDA, 2010A. FDA BRIEFING DOCUMENT: ADVISORY COMMITTEE MEETING FOR NDA 21071 AVANDIA (ROSIGLITAZONE MALEATE) TABLET (JULY 13 AND 14, 2010). [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DOWNLOADS/ADVISORYCOMMITTEES/COMMITTEESMEETINGMATERIALS/DRUGS/ENDOCRINOLOGICANDMETABOLICDRUGSADVISORYCOMMITTEE/UCM218493.PDF [ACCESSED 01 APRIL 2015].

FDA, 2011. FDA DRUG SAFETY COMMUNICATION: UPDATED RISK EVALUATION AND MITIGATION STRATEGY (REMS) TO RESTRICT ACCESS TO ROSIGLITAZONE CONTAINING MEDICINES INCLUDING AVANDIA, AVANDAMET, AND AVANDARYL [PRESS RELEASE: 18-05-2011]. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DRUGS/DRUGSAFETY/UCM255005.HTM#SAFETY_ANNOUNCEMENT [ACCESSED 23 APRIL 2015].

FDA, 2013A. STRUCTURED APPROACH TO BENEFIT-RISK ASSESSMENT IN DRUG REGULATORY DECISION-MAKING: DRAFT PATIENT DRUG USER FREE ACT (PDUFA) I IMPLEMENTATION PLAN. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DOWNLOADS/FORINDUSTRY/USERFEES/PRESCRIPTIONDRUGUSERFEE/UCM329758.PDF [ACCESSED 02 APRIL 2015].

FDA, 2013B. READJUDICATION OF THE ROSIGLITAZONE EVALUATED FOR CV OUTCOMES AND REGULATION OF GLYCEMIA IN DIABETES TRIAL (RECORD). JOINT MEETING OF THE ENDOCRINOLOGIC AND METABOLIC DRUGS ADVISORY COMMITTEE AND THE DRUG SAFETY AND RISK MANAGEMENT ADVISORY COMMITTEE. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DOWNLOADS/ADVISORYCOMMITTEES/COMMITTEESMEETINGMATERIALS/DRUGS/ENDOCRINOLOGICANDMETABOLICDRUGSADVISORYCOMMITTEE/UCM354859.PDF [ACCESSED 02 APRIL 2015].

FDA, 2013C. FDA DRUG SAFETY COMMUNICATION: FDA REQUIRES REMOVAL OF SOME PRESCRIBING AND DISPENSING RESTRICTIONS FOR ROSIGLITAZONE-CONTAINING DIABETES MEDICINES [PRESS RELEASE: 25-11-2013]. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DRUGS/DRUGSAFETY/UCM376389.HTM [ACCESSED 23 APRIL 2015].

GLASZIOU, P., VANDENBROUCKE, J. & CHALMERS, I., 2004. ASSESSING THE QUALITY OF RESEARCH. BMJ, 328(7430), PP. 39-41.

GUO, J. ET AL., 2010. A REVIEW OF QUANTITATIVE RISK-BENEFIT METHODOLOGIES FOR ASSESSING DRUG SAFETY AND EFFICACY-REPORT OF THE ISPOR RISK-BENEFIT MANAGEMENT WORKING GROUP. VALUE HEALTH, 13(5), PP. 657-666.

HALLGREEN, C., ET AL., 2014. BENEFIT-RISK ASSESSMENT IN A POST-MARKET SETTING: A CASE STUDY INTEGRATING REAL-LIFE EXPERIENCE INTO BENEFIT-RISK METHODOLOGY. PHARMACOEPIDEMIOL DRUG SAF, 23(9), PP. 974-983.

HAMMAD, T. ET AL., 2013. THE FUTURE OF POPULATION-BASED POSTMARKET DRUG RISK ASSESSMENT: A REGULATOR'S PERSPECTIVE. CLIN PHARMACOL THER, 94(3), PP. 349-358.

HIGGINS, J., THOMPSON, S., DEEKS, J. & ALTMAN, D., 2003. MEASURING INCONSISTENCY IN META-ANALYSES. BMJ, 327(7414), PP. 557-560.

HOLDEN, W., JUHAERI, J. & DAI, W., 2003. BENEFIT-RISK ANALYSIS: A PROPOSAL USING QUANTITATIVE METHODS. PHARMACOEPIDEMIOL DRUG SAF, 12(7), PP. 611-616.

HOLDEN, W., JUHAERI, J. & DAI, W., 2003. BENEFIT-RISK ANALYSIS: EXAMPLES USING QUANTITATIVE METHODS. PHARMACOEPIDEMIOL DRUG SAF, 12(8), PP. 693-697.

HOME, P. ET AL., 2009. ROSIGLITAZONE EVALUATED FOR CARDIOVASCULAR OUTCOMES IN ORAL AGENT COMBINATION THERAPY FOR TYPE 2 DIABETES (RECORD): A MULTICENTRE, RANDOMISED, OPEN-LABEL TRIAL. LANCET, 373(9681), PP. 2125-2135.

KRALL, R., 2007. CARDIOVASCULAR SAFETY OF ROSIGLITAZONE. LANCET, 369(9578), PP. 1995-1996.

LAUPACIS, A., SACKETT, D. & ROBERTS, R., 1988. AN ASSESSMENT OF CLINICALLY USEFUL MEASURES OF THE CONSEQUENCES OF TREATMENT. N ENGL J MED, 318(26), PP. 1728-1733.

LESTER, J. ET AL., 2013. EVALUATION OF FDA SAFETY-RELATED DRUG LABEL CHANGES IN 2010. PHARMACOEPIDEMIOL DRUG SAF, 22(3), PP. 302-305.

LOKE, Y., KWOK, C. & SINGH, S., 2011. COMPARATIVE CARDIOVASCULAR EFFECTS OF THIAZOLIDINEDIONES: SYSTEMATIC REVIEW AND META-ANALYSIS OF OBSERVATIONAL STUDIES. BMJ, VOLUME 342, P. D1309.

LOOK AHEAD RESEARCH GROUP, 2013. CARDIOVASCULAR EFFECTS OF INTENSIVE LIFESTYLE INTERVENTION IN TYPE 2 DIABETES. N ENGL J MED, 369(2), PP. 145-154.

MARX, A. & BUCHER, H., 2003. NUMBERS NEEDED TO TREAT DERIVED FROM META-ANALYSIS: A WORD OF CAUTION. ACP J CLUB, 138(2), PP. A11-A12.

MCALISTER, F., 2002. COMMENTARY: RELATIVE TREATMENT EFFECTS ARE CONSISTENT ACROSS THE SPECTRUM OF UNDERLYING RISKS...USUALLY. INT J EPIDEMIOL, 31(1), PP. 76-77.

MCNAUGHTON, R., HUET, G. & SHAKIR, S., 2014. AN INVESTIGATION INTO DRUG PRODUCTS WITHDRAWN FROM THE EU MARKET BETWEEN 2002 AND 2011 FOR SAFETY REASONS AND THE EVIDENCE USED TO SUPPORT THE DECISION-MAKING. BMJ OPEN, 4(1), P. E004221.

MT-ISA, S. ET AL., 2014. BALANCING BENEFIT AND RISK OF MEDICINES: A SYSTEMATIC REVIEW AND CLASSIFICATION OF AVAILABLE METHODOLOGIES. PHARMACOEPIDEMIOL DRUG SAF, 23(7), PP. 667-678.

MT-ISA, S. ET AL., 2013. REVIEW OF METHODOLOGIES FOR BENEFIT AND RISK ASSESSMENT OF MEDICATION, LONDON, UK: PROTECT CONSORTIUM.

NISSEN, S. & WOLSKI, K., 2007. EFFECT OF ROSIGLITAZONE ON THE RISK OF MYOCARDIAL INFARCTION AND DEATH FROM CARDIOVASCULAR CAUSES. N ENGL J MED, 356(24), PP. 2457-2471.

NISSEN, S. & WOLSKI, K., 2010. ROSIGLITAZONE REVISITED: AN UPDATED META-ANALYSIS OF RISK FOR MYOCARDIAL INFARCTION AND CARDIOVASCULAR MORTALITY. ARCH INTERN MED, 170(14), PP. 1191-1201.

NIXON, R. ET AL., 2016. A CASE STUDY USING THE PROACT-URL AND BRAT FRAMEWORKS FOR STRUCTURED BENEFIT RISK ASSESSMENT. BIOM J, 58(1), PP. 8-27.

PAPANIKOLAOU, P., CHRISTIDI, G. & IOANNIDIS, J., 2006. COMPARISON OF EVIDENCE ON HARMS OF MEDICAL INTERVENTIONS IN RANDOMIZED AND NONRANDOMIZED STUDIESCOMPARISON OF EVIDENCE ON HARMS OF MEDICAL INTERVENTIONS IN RANDOMIZED AND NONRANDOMIZED STUDIES. CMAJ, 174(5), PP. 635-641.

PSATY, B. & FURBERG, C., 2007. ROSIGLITAZONE AND CARDIOVASCULAR RISK. N ENGL J MED, 356(24), PP. 2522-2524.

QURESHI, Z. ET AL., 2011. MARKET WITHDRAWAL OF NEW MOLECULAR ENTITIES APPROVED IN THE UNITED STATES FROM 1980 TO 2009. PHARMACOEPIDEMIOL DRUG SAF, 20(7), PP. 772-777.

SHUSTER, J. & SCHATZ, D., 2008. THE ROSIGLIAZONE META-ANALYSIS: LESSONS FOR THE FUTURE. DIABETES CARE, 31(3), P. E10.

STRAUS, S., GLASZIOU, P., RICHARDSON, W. & HAYNES, R., 2011. EVIDENCE-BASED MEDICINE: HOW TO PRACTICE AND TEACH IT. 4TH ED. LONDON, UK: CHURCHILL LIVINGSTONE.

VAN STAA, T. ET AL., 2008. EVALUATING DRUG TOXICITY SIGNALS: IS A HIERARCHICAL CLASSIFICATION OF EVIDENCE USEFUL OR A HINDRANCE? PHARMACOEPIDEMIOL DRUG SAF, 17(5), PP. 475-484.

VANDENBROUCKE, J., 2004. WHEN ARE OBSERVATIONAL STUDIES AS CREDIBLE AS RANDOMISED TRIALS? LANCET, 363(9422), PP. 1728-1731.

VANDENBROUCKE, J., 2006. WHAT IS THE BEST EVIDENCE FOR DETERMINING HARMS OF MEDICAL TREATMENT? CMAJ, 174(5), PP. 645-646.

VANDENBROUCKE, J. & PSATY, B., 2008. BENEFITS AND RISKS OF DRUG TREATMENTS: HOW TO COMBINE THE BEST EVIDENCE ON BENEFITS WITH THE BEST DATA ABOUT ADVERSE EFFECTS. JAMA, 300(20), PP. 2417-2419.

WALKER, S. ET AL., 2015. A UNIVERSAL FRAMEWORK FOR THE BENEFIT-RISK ASSESSMENT OF MEDICINES: IS THIS THE WAY FORWARD? THERAPEUTIC INNOVATION REGULATORY SCI, 49(1), PP. 17-25.

WYSOWSKI, D. & SWARTZ, L., 2005. ADVERSE DRUG EVENT SURVEILLANCE AND DRUG WITHDRAWALS IN THE UNITED STATES, 1969-2002: THE IMPORTANCE OF REPORTING SUSPECTED REACTIONS. ARCH INTERN MED, 165(12), PP. 1363-1369.

YUAN, Z., LEVITAN, B. & BERLIN, J., 2011. BENEFIT-RISK ASSESSMENT: TO QUANTIFY OR NOT TO QUANTIFY, THAT IS THE QUESTION. PHARMACOEPIDEMIOL DRUG SAF, 20(6), PP. 653-656.

II.7. SUPPLEMENTAL DATA II

II.7.1. SUPPLEMENTAL DATA II.1 - SEARCH STRATEGY

Supplemental Table II. 1 – Search strategies used to identify RCTs (A) and observational studies (B) aimed to evaluate cardiovascular adverse events associated with the use of rosiglitazone and pioglitazone.

Search strategy A
1. "Rosiglitazone OR Pioglitazone" 2. Filters activated: "Clinical Trial", "Clinical Trial, Phase IV", "Randomized Controlled Trial", "Clinical Trial, Phase III", "Clinical Trial, Phase II", and "Humans" 3. Limit 2 to English language
Search strategy B
1. "Rosiglitazone OR Avandia OR pioglitazone OR Actos OR Thiazolidinedione OR Thiazolidinediones OR TZD OR TZDs" 2. "Cohort OR case control OR case-control OR observational OR epidemiologic OR retrospective OR meta analysis OR meta-analysis OR meta analyses OR meta-analyses" 3. "Cardiovascular OR cardiac OR coronary OR ischemic OR ischemia OR myocardial OR revascularization OR heart OR CVD OR CAD OR IHD OR HF OR CHF OR hospital OR mortality OR death OR stroke OR cerebrovascular accident OR CVA OR cerebral haemorrhage OR subarachnoid haemorrhage OR cerebral thrombosis OR cerebral infarction OR brain infarction OR cerebral infarct" 4. 1 AND 2 AND 3 5. Limit 4 to English language

II.7.2. SUPPLEMENTAL DATA II.2 - CHARACTERISTICS OF THE OBSERVATIONAL STUDIES

Supplemental Table II. 2 – Characteristics of the observational studies included in the meta-analyses.

Reference	Design	Outcomes Evaluated	Population Details	N=
Azoulay, et al. 2009	Nested case control study	First stroke	The cohort comprised 75,717 patients over the age of 40 who were prescribed a first OHA, of whom 2,417 had a stroke during follow-up. Up to 10 controls were matched to each case on age, sex, date of cohort entry, and duration of follow- up. Subjects who initiated their treatment with insulin were excluded. Mean age for cases and controls 74.1 and 73.8 years, respectively.	Cases, n=2,416; Controls, n=23,987
Bilik, et al. 2010	Retrospective cohort study	NMI, CRV, Nonfatal stroke; CV death, all-cause death	Type 2 diabetes patients (by prescription); exclude age at diagnosis <30 years and treatment with insulin only.	Any TZD prescription, n=1,815; RSG alone, n=773; PIO alone, n=711; multiple TZDs, n=331
Breunig, et al. 2014	Retrospective cohort study	CHF	The study population included beneficiaries, between the ages of 18 and 64, with at least 1 diagnosis of type 2 diabetes, who were started on MET monotherapy or any drug containing PIO or RSG and had no history of MET or TZD use in the prior 6 months.	MET, n=5,548; PIO, n=413; RSG, n=310
Brownstein, et al. 2010	Retrospective cohort study	AMI	DM patients >18 years; ICD9 code DM 250.XX or an AIC>6% and ≥1 record of prescription.	RSG, n=1,879; MET, n=12,490; SU, n=11,200; PIO, n=806
Chou, et al. 2011	Retrospective cohort study (with secondary data analysis)	MI, CHF, angina, stroke	Type 2 DM Taiwanese patients (by prescription and ICD-9 diagnosis codes).	PIO, n=1,677; RSG, n=6,048
Dore, et al. 2009	Nested case-control study	AMI	Base cohort of 307,121 patients from 5 states Medicaid claims, making source population 95,332 individuals who used MET plus SU. For 2316 cases, 9700 controls were randomly selected matched with age- and state of residence. More than 40% of participants were aged 70 years or older.	Cases, n=2,316; Controls, n=9,700
Dormuth, et al. 2009	Nested case control study	AMI	158,578 patients with Type 2 diabetes who used MET as first-line drug treatment; 2,244 AMI cases and 8,903 matched controls. Mean age for cases and controls 70 years.	Cases, n=2,244; Controls, n=8,903
Gallagher, et al. 2011	Retrospective cohort study	ACS, stroke, CHF; all-cause death (including cause of death)	Type 2 DM patients age 40 years and over (from the UK GPRD). (Note: Study patients may have received other antidiabetic medications but it is not clear in the publication.)	MET, n=121,637; SU, n=76,863; RSG, n=22,636; PIO, n=18,953; insulin, n=26,458
Gerrits, et al. 2007	Retrospective cohort study	AMI, CRV	All patients with a diagnostic code [ICD-9: 250.xx] were initially extracted. Exclusion criteria: dispensed both PIO and RSG, unknown gender, gaps in their insurance coverage; younger than 45 years of age, had less than 6 months of history in the database, and had been dispensed less than two prescriptions of the index TZD within 6 months after the index date were excluded. Mean age were 58 for both Rosi and Pio cohorts.	PIO, n=14,807; RSG, n=15,104
Graham, et al. 2010	Retrospective cohort study	AMI, stroke, CHF; all-cause death	TZD-exposed patients 65 years and older. (NOTE: Some patients received other antidiabetic medication.)	RSG, n=67,593; PIO, n=159,978
Habib, et al. 2009	Retrospective cohort study	Fatal and non-fatal AMI (primary), hospitalization for CHF, fatal and non-fatal stroke, TIA, CHD, all-cause death	All patients had prescription coverage, >18 years; at least one clinical encounter with a coded diagnosis of diabetes and at least one prescription of an oral diabetes medication; at least 12 months of continuous enrolment in the HMO prior to the index date, and at least 6 months of follow-up after the index date. Mean age was 58 years for the cohort; 59 and 57 for RSG and PIO cohorts, respectively.	RSG alone, n=1,056; PIO alone, n=3,217; both RSG and PIO, n=307
Hsiao, et al. 2009	Retrospective Cohort Study	MI, CHF, AP, stroke, TIA, and composite of any of these outcomes	Newly diagnosed patients with T2DM (ICD-9: 250.xx) and were prescribed oral anti-hyperglycaemic agents (SU, MET and/or a TZD) at least three times between 03/01/2001 and 12/31/2005 (n = 473 483). None of these patients	SU + MET based therapy, n=317,246; SU based therapy, n=104,023; MET based therapy, n=49,626;

Reference	Design	Outcomes Evaluated	Population Details	N=
			had records showing a diagnosis of diabetes during the year before the index date. Mean age were 61 2 and 60.8 for RSG and PIO cohorts, respectively.	RSG alone, n=2,093; PIO alone, n=495
Juurlink, et al. 2009	Retrospective cohort study	Composite of death or hospital admission for AMI or CHF	Patient characteristics, proportion of prior cardiovascular admissions and procedures, history of medications were highly similar for the two drug groups; patients aged 66 years or older; 69.1% and 68.7% patients 66-75 years for RSG and PIO cohorts, respectively.	PIO, n=16,951; PIO, n=22,785
Karter, et al. 2005	Retrospective cohort study	CHF	All patients in the Kaiser Permanente Medical Care Program with Type 2 diabetes (23,440) between Oct 1999 and Nov 2001. Only patients initiating single new therapies were included. Mean age was 59 years for the cohort; 60 and 59 for PIO and MET cohorts, respectively	PIO, n=3,556; SU, n=5,921; MET, n=11,937; insulin, n=2,026
Koro, et al. 2008	Nested case control study	MI	IHCIS contains total of 891,901 base diabetic non-elderly, insurance-carrying population in the USA, mean age was 63 years for the cases and controls.	Cases, n=9,870; Controls, n=29,610
Lipscombe, et al. 2007	Nested case-control	CHF, AMI, all-cause death	Ontarians aged 66 years or older with diabetes as identified in the Ontario Diabetes Database and who were dispensed at least 1 oral hypoglycaemic agent. For each case, up to 5 controls were randomly selected and matched on age (±1 year), sex, diabetes duration (2 years, 2-5 years, or 5 years), and history of CVD within 5 years of cohort entry. In the CHF and AMI analyses, controls were also matched on history of an event (within 1 year of cohort entry and within 1-5 years). For different outcomes, mean age for cases and controls were 76.5-78.6 and 76.4-78.7 years.	CHF: cases, n= 12,491; controls, n=61,827; Acute MI: cases, n=12,578; controls, n=62,651; All-cause death: cases, n=30,265; controls, n=150,650
Loebstein, et al. 2011	Retrospective cohort study	AMI, ACS, CRV, CHF, all-cause death	Candidates were drawn from the Maccabi diabetes mellitus registry, which includes all patients defined by the American Diabetes Association criteria. In addition, the registry includes patients dispensed hypoglycaemic medications or at least one HbA1C level ≥7.25%. Study patients were defined as those who purchased the RSG and/or MET for a period of at least 6 months.	RSG alone, n=745; RSG + MET, n=2,753; MET alone, n=11,938
Lu, et al. 2013	Retrospective cohort study	Acute ischemic stroke, acute intracerebral haemorrhage, CHF and AMI	Eligible study subjects were patients aged 18 years and older with a diagnosis of type 2 diabetes (ICD-9-CM code 250.x0 or 250.x2) on outpatient claims and/or hospitalization records. Subjects who were diagnosed with major macrovascular events, including stroke (ICD-9-CM codes 430–438), AMI (AMI; ICD-9-CM codes 410–411), old MI (ICD-9-CM code 412) or CHF (ICD-9-CM code 428), prior to their index date were excluded.	non-TZDs, n=10,316; RSG, n=2,996; PIO, n=2,669
Margolis, et al. 2008	Retrospective cohort study	Any serious atherosclerotic vascular disease of the heart (includes MI, unstable angina, CV death, CARP)	All subjects enrolled in this study were required to have at least two records for diabetes between January 2002 and 2006 and ≥ 40 years old. The database diagnosis of diabetes was previously validated. First study, all diabetics could have been diagnosed with diabetes at any time since they had been enrolled. An individual could have had drug exposures or an outcome before 2002, but not contribute to our study. Second study, a smaller sub-cohort, patients' first THIN diagnosis for diabetes and first drug treatment for diabetes must both have occurred after January 2002. There were 35% and 41% patients aged 70 years and older for all diabetics and new onset diabetics, respectively.	insulin, n=16,213; SUs, n=32,857; biguanide, n=43801; meglitinide, n=1,061; RSG, n=7282; PIO, n=2244
McAfee, et al. 2007	Retrospective cohort study	MI, CRV, CPT	All initiators of RSG, MET, and SU for whom the first recorded dispensing followed 1) at least six months' membership; and 2) the member's 18th birthday. Patients were required to have medical and pharmacy benefits. Three study groups: monotherapy and dual-therapy, and combination with insulin. Mean age were 51-52 years for RSG, or MET, or SU, either monotherapy, or dual-therapy groups, or combination-with-insulin group.	RSG, n=8,977; MET, n=8,977; SU, n=8,977
Morgan, et al. 2012	Retrospective cohort study	All-cause death, major adverse cardiovascular events (MACE),	Primary care patients with type 2 diabetes who had MET monotherapy as their first treatment and who then initiated on relevant second-line, glucose-lowering regimens	MET + SU, n=15,377; SU, n=2,244; MET + PIO, n=2,525; MET + RSG, n=4,677; MET +

Reference	Design	Outcomes Evaluated	Population Details	N=
		cancer, and a combined end point of any of these		DPP4, n=1,455
Pantalone, et al. 2009	Retrospective cohort study	CAD, CHF, death	Type 2 diabetes with prescription for RSG, PIO, MET, or SU, age >18 years with no history of dialysis, CAD, or HF; excluded if prescribed insulin or multiple oral agents	rosiglitazone, n=1,079; MET, n=10,436; SU, n=7,427; PIO, n=1,508
Rajagopalan, et al. 2004	Retrospective cohort study	CHF	Patients aged >18 years with a diagnosis of type 2 diabetes and/or evidence of use of antidiabetic medications who began receiving PIO or insulin. Exclusion criteria: patients who had a prior diagnosis of CHF or used digoxin, who used troglitazone at any time, or who used any oral antidiabetic drug other than MET or a SU during the final 6 months of the pre-index period. Patients who had facility or provider claims with a diagnosis of CHF at any time during the pre-index period and those who were not eligible for health and pharmacy benefits during the entire pre-treatment and follow-up periods were also excluded. No. of patients before matching for PIO and Insulin are 3870 and 2577. Mean age was 51 years for the cohort; 51 for both Pio and Insulin cohorts.	PIO, n=1,668; insulin, n=1,668
Ramirez, et al. 2009	Prospective cohort study	CV death, all-cause death	Patients with diabetes and chronic kidney disease; USA population of the Dialysis Outcomes and Practice Patterns Study (DOPPS)	RSG, n=177; PIO, n=118; non-TZDs, n=2,050
Stockl, et al. 2009	Nested case-control study	AMI	Risk of AMI with RSG or PIO exposure compared to no TZD exposure; base cohort of 230,858 patients with OHA or exenatide prescription in a large USA PBM. Total 1681 AMI cases were identified and matched with 6653 controls. Mean age for cases and controls were both 73 years.	Cases, n=1,681; Controls, n=6,653
Tannen, et al. 2012	Retrospective cohort study	MI, CHF, all-cause death	Patients aged 35-90 years identified oral antidiabetic treatments of individual patients from prescription records. Mean age was 65 years for the cohort; 64.5 and 64.8 for RSG and PIO cohorts, respectively.	Replication studies: PIO, n=709; non-PIO, n=1,654; RSG, n=2,001; non-RSG, n=5,056
Tzoulaki, et al. 2009	Retrospective cohort study	MI, CHF, and all-cause death	Patients (men and women with diabetes included in the general practice research database in the United Kingdom) aged 35-90 years with an episode of care between 1 January 1990 and 31 December 2005 and a diagnostic (Read) code associated with a clinical or referral event for diabetes. Records with multiple or missing date of death were excluded.	SUs, n=64,148; RSG, n=8,442; RSG combination, n=9,640; PIO monotherapy or combination, n=3,816; other drug combinations, n=37,253; MET, n=68,181
Vallarino, et al. 2013	Retrospective cohort study	Composite of MI or stroke requiring hospitalization	The study population included patients with T2DM C 45 years old who were new users of either PIO or insulin	pioglitazone, n=38,588; insulin, n=17,948
Vanasse, et al. 2009	Nested case-control study	All-cause or CV death, AMI, CHF, stroke	All diabetic patients aged 65 years or older living in the province of Québec between January 2001 and December 2002. Mean age for cases and controls = 75.6 and 75.1 years.	All-cause death: cases, n=18,554; controls, n=370,866; CV death: cases, n=4,455; controls, n=89,037
Walker, et al. 2008	Retrospective cohort study	MI, CRV	All users of RSG, PIO, MET, and a SU for whom the first recorded dispensing; followed (1) ≥ 6 months membership; and (2) ≥ 18 year-old. Patients were required to have medical and pharmacy benefits. Little information on persons over the age of 65. Regimens involving RSG and PIO were more similar to one another in patient characteristics than were other regimens, although PIO-using groups in general had a higher prevalence of baseline dyslipidaemia than did RSG-using groups.	RSG n=57,000; PIO, n=51,000; MET, n=275,000; SU, n=160,000
Wheeler, et al. 2013	Retrospective cohort study	All-cause death	New users of oral hypoglycaemic medication monotherapy between 2004 and 2009 who received care for at least 1 year from the Veterans Health Administration	MET, n=132,306; glipizide, n=28,957; glibenclamide, n=28,156; RSG, n=3,753
Wertz, et al. 2010	Retrospective cohort study	AMI, AHF, All-cause death	Patients 18 years of age with a new RSG or PIO claim between January 1, 2001, and December 31,	RSG, n=18,319; PIO, n=18,309

Reference	Design	Outcomes Evaluated	Population Details	N=
			2005	
Winkelmayer, et al. 2008	Retrospective cohort study	All-cause death (primary), MI, stroke, CHF	New RSG or PIO users (≥6months) of US Medicare beneficiaries older than 65 years (N=28,361). Patient characteristics, proportion of prior cardiovascular procedures and medications are comparable for the two drug groups. Mean age were 76.3 for both RSG and PIO cohorts.	PIO, n=14,260; RSG, n=14,101
Yang, et al. 2014	Retrospective cohort study	All-cause death	The study population included patients with type 2 diabetes mellitus but without type 1 diabetes, gestational diabetes, diabetes insipidus, and renal glycosuria prior to the index date (the first claim date of PIO or INS between 1 January 2003, and 31 December 2008), aged 45 years, with a baseline 6 months and follow-up 1 month, new users of PIO or INS, and two claims of PIO or INS, respectively, on or within 6 months after the index date	PIO, n=38,588; insulin, n=17,948
Ziyadeh, et al. 2009	Retrospective cohort study	MI, CR, all-cause death	i3 Drug Safety has access to a proprietary integrated research database of health insurance plan members who have both medical and prescription drug benefits. There were 57.6% and 57.3% patients younger than 55 years for RSG and PIO cohorts, respectively.	RSG, n=47,501; PIO, n=47,501

ACS, acute coronary syndrome; AHF, acute heart failure; AMI, acute myocardial infarction; CAD, coronary artery disease; CHD, coronary heart disease; CHF, congestive heart failure; CRV, coronary revascularization; CV, cardiovascular; CVD, cardiovascular disease; MET, metformin; MI, myocardial infarction; NMI, nonfatal MI; OHA, oral hypoglycaemic agent; PIO, pioglitazone; RSG, rosiglitazone; SU, sulfonylurea; TIA, transient ischemic attack; TZD, thiazolidinedione; USA, United States of America.

II.7.3. SUPPLEMENTAL DATA II.3 – ROSIGLITAZONE IN OBSERVATIONAL STUDIES

Supplemental Table II. 3 – Adjusted OR (95%CI), p-value and I^2 for CV adverse events associated with the use of rosiglitazone in observational studies (meta-analyses).

Studies controlled with placebo												
CV event	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2
MIF	1	1.14 (0.90, 1.44)	0.267	NA	2	0.86 (0.64, 1.17)	0.349	0%	3	1.01 (0.77, 1.32)	0.935	9.40%
Stroke	0	NA	NA	NA	0	NA	NA	NA	0	NA	NA	NA
CHF	0	NA	NA	NA	2	1.25 (0.99, 1.58)	0.056	0%	2	1.25 (0.99, 1.58)	0.056	0%
CVD	0	NA	NA	NA	0	NA	NA	NA	0	NA	NA	NA
All-cause death	0	NA	NA	NA	2	0.88 (0.75, 1.05)	0.167	0%	2	0.88 (0.75, 1.05)	0.167	0%
Studies controlled with active therapy												
CV event	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2
MIF	7	1.13 (0.98, 1.31)	0.099	74.23%	17	**1.22 (1.04, 1.44)**	**0.015**	70.29%	24	**1.18 (1.06, 1.32)**	0.002	70.51%
Stroke	3	1.03 (0.82, 1.30)	0.805	33.39%	6	1.08 (0.81, 1.46)	0.594	72.03%	9	1.05 (0.88, 1.27)	0.574	63.42%
CHF	3	**1.74 (1.37, 2.20)**	**<0.001**	83.03%	11	**1.32 (1.11, 1.56)**	**0.001**	54.53%	14	**1.43 (1.23, 1.65)**	<0.001	71.35%
CVD	1	0.88 (0.69, 1.12)	0.301	NA	2	1.34 (0.98, 1.84)	0.069	49.60%	3	1.15 (0.82, 1.62)	0.405	75.12%
All-cause death	3	1.16 (0.86, 1.56)	0.346	90.60%	9	1.06 (0.88, 1.30)	0.535	79.87%	12	1.09 (0.94, 1.27)	0.269	82.15%
Monotherapy studies												
CV event	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2
MIF	2	1.56 (0.99, 2.44)	0.053	30.80%	11	**1.26 (1.01, 1.57)**	**0.053**	74.10%	13	**1.31 (1.08, 1.60)**	**0.007**	72.10%
Stroke	2	1.14 (0.98, 1.33)	0.091	0%	3	1.31 (0.94, 1.83)	0.111	0%	5	**1.17 (1.02, 1.35)**	**0.027**	0%
CHF	2	**1.96 (1.41, 2.72)**	**<0.001**	**0%**	8	**1.25 (1.02, 1.54)**	**0.032**	61.30%	10	1.54 (0.99, 2.37)	0.054	76.30%
CVD	1	0.88 (0.59, 1.31)	0.529	NA	2	1.34 (0.98, 1.84)	0.069	49.60%	3	1.11 (0.73, 1.67)	0.631	75.10%
All-cause death	2	1.11 (0.71, 1.74)	0.648	91.30%	7	1.13 (0.87, 1.46)	0.355	78.40%	9	1.12 (0.90, 1.41)	0.303	86.20%
Add-on studies												
CV event	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2
MIF	6	1.06 (0.97, 1.15)	0.204	17.66%	8	1.09 (0.91, 1.32)	0.347	56.40%	14	1.06 (0.97, 1.16)	0.23	41.32%
Stroke	1	0.81 (0.59, 1.12)	0.198	NA	3	0.95 (0.60, 1.50)	0.8	87.93%	4	0.91 (0.64, 1.31)	0.621	84.75%
CHF	1	**1.43 (1.25, 1.63)**	**<0.001**	NA	5	**1.36 (1.18, 1.56)**	**<0.001**	2.03%	6	**1.39 (1.27, 1.53)**	**<0.001**	0%
CVD	0	-	-	-	0	-	-	-	0	-	-	-
All-cause death	1	**1.26 (1.12, 1.42)**	**<0.001**	0%	4	**0.89 (0.80, 0.98)†**	**0.021**	**0%**	5	0.96 (0.80, 1.16)	0.695	79.11%
Overall												
CV event	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2
MIF	8	1.13 (0.99, 1.29)	0.059	69.94%	18	**1.12 (1.04, 1.42)**	0.015	68.34%	26	**1.17 (1.06, 1.30)**	0.002	67.85%
Stroke	3	1.03 (0.82, 1.30)	0.805	44.39%	5	1.17 (0.84, 1.62)	0.365	69.01%	8	1.10 (0.90, 1.33)	0.355	61.27%
CHF	3	**1.74 (1.37, 2.20)**	<0.001	83.03%	13	**1.31 (1.14, 1.51)**	<0.001	46.41%	16	**1.41 (1.23 ,1.61)**	<0.001	69.09%
CVD	1	0.88 (0.69, 1.12)	0.301	0%	2	1.34 (0.98, 1.84)	0.069	49.60%	3	1.15 (0.82, 1.62)	0.405	75.12%
All-cause death	3	1.16 (0.86, 1.56)	0.346	90.62%	10	1.04 (0.87, 1.25)	0.652	78.32%	13	1.07 (0.93, 1.24)	0.351	81.22%

CHF, congestive heart failure; CV, cardiovascular; MI, myocardial infarction; n, number of studies (one publication could provide more than one estimate and each estimate was considered one study for the total n) OR, odds ratio.
Bold values are statistically significant (95% Confidence Interval).
† Protective effect.

II.7.4. SUPPLEMENTAL DATA II.4 – PIOGLITAZONE IN OBSERVATIONAL STUDIES

Supplemental Table II. 4 – Adjusted OR (95% CI), p-value and I^2 for CV adverse events associated with the use of pioglitazone in observational studies (meta-analyses).

Studies controlled with placebo												
CV events	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2
MI	1	1.21 (0.87, 1.67)	0.25	-	1	0.71 (0.39, 1.30)	0.265	-	2	0.99 (0.59, 1.64)	0.959	57.10%
Stroke	0	-	-	-	0	-	-	-	0	-	-	-
CHF	0	-	-	-	1	1.25 (0.99, 1.58)	0.46	-	1	1.25 (0.99, 1.58)	0.46	-
CV death	0	-	-	-	0	-	-	-	0	-	-	-
All-cause death	0	-	-	-	1	**0.69 (0.49, 0.98)**	0.036	-	1	**0.69 (0.49, 0.98)**	0.036	-
Studies controlled with active therapy												
CV events	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2
MI	5	0.96 (0.80, 1.15)	0.651	54.03%	8	0.80 (0.60, 1.05)	0.109	77.98%	13	0.85 (0.69, 1.04)	0.116	84.14%
Stroke	1	0.89 (0.49, 1.60)	0.185	-	3	0.88 (0.75, 1.04)	0.16	35.29%	4	0.88 (0.75, 1.03)	0.069	17.27%
CHF	1	1.07 (0.89, 1.28)	0.378	-	7	0.92 (0.66, 1.29)	0.585	91.43%	8	0.93 (0.70, 1.23)	0.521	92.45%
CV death	0	-	-	-	1	1.37 (0.77, 2.43)	0.281	-	1	1.37 (0.77, 2.43)	0.281	-
All-cause death	2	1.15 (0.96, 1.38)	0.14	0.24%	7	0.69 (0.44, 1.10)	0.121	95.62%	9	0.77 (0.49, 1.20)	0.245	96.77%
Monotherapy studies												
CV events	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2
MI	1	0.73 (0.50, 1.35)	0.313	-	4	**0.84 (0.71, 0.99)†**	**0.047**	0%	5	**0.83 (0.71, 0.98)†**	**0.029**	0%
Stroke	1	1.25 (0.61, 2.55)	0.539	-	3	0.92 (0.72, 1.19)	0.519	0%	5	0.95 (0.75, 1.21)	0.688	0%
CHF	1	0.91 (0.52, 1.59)	0.786	-	6	1.10 (0.86, 1.39)	0.456	51.90%	7	1.07 (0.86, 1.35)	0.539	45.10%
CV death	0	-	-	-	1	1.37 (0.77, 2.43)	0.281	-	1	1.37 (0.77, 2.43)	0.281	-
All-cause death	1	0.94 (0.44, 2.00)	0.872	-	4	0.82 (0.57, 1.16)	0.26	75.30%	5	0.84 (0.61, 1.15)	0.277	67.90%
Add-on studies												
CV events	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2
MI	5	1.02 (0.87, 1.19)	0.828	50.15%	5	0.76 (0.50, 1.14)	0.18	82.28%	10	0.87 (0.68, 1.12)	0.294	88.47%
Stroke	1	0.68 (0.38, 1.20)	0.185	-	3	0.80 (0.59, 1.09)	0.16	35.29%	4	0.80 (0.63, 1.02)	0.069	17.27%
CHF	1	1.09 (0.90, 1.32)	0.378	-	7	0.84 (0.52, 1.36)	0.484	91.43%	8	0.87 (0.58, 1.32)	0.521	92.45%
CV death	0	-	-	-	0	-	-	-	0	-	-	-
All-cause death	1	1.20 (0.98, 1.47)	0.078	-	4	0.59 (0.34, 1.03)	0.065	95.48%	5	0.69 (0.37, 1.29)	0.243	97.80%
Overall												
CV events	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2
MIF	6	0.99 (0.86, 1.16)	0.986	47.03%	8	0.79 (0.59, 1.05)	0.099	77.85%	14	0.87 (0,71, 1,06)	0,162	83,63%
Stroke	2	0.89 (0.49, 1.60)	0.692	41.59%	5	0.92 (0.77 - 1.09)	0.325	0%	7	0.91 (0.76, 1.07)	0.249	0%
CHF	2	1.07 (0.89, 1.28)	0.467	0%	13	0.94 (0.68, 1.28)	0.679	90.03%	15	0.94 (0.72, 1.23)	0.669	88.96%
CVD	0	-	-	-	1	1.37 (0.77, 2.43)	0.281	0%	1	1.37 (0.77, 2.43)	0.281	0%
All-cause death	2	1.15 (0.96, 1.38)	0.141	0.24%	7	0.69 (0.44, 1.10)	0.119	95.53%	9	0.76 (0.48, 1.20)	0.243	96.74%

CHF, congestive heart failure; CV, cardiovascular; MI, myocardial infarction; n, number of studies (one publication could provide more than one estimate and each estimate was considered one study for the total n) OR, odds ratio.
Bold values are statistically significant (95% Confidence Interval).
† Protective effect.

II.7.5. SUPPLEMENTAL DATA II.5 – ROSIGLITAZONE VERSUS PIOGLITAZONE IN OBSERVATIONAL STUDIES

Supplemental Table II. 5 – Adjusted OR (95% CI), p-value and I^2 for CV adverse events associated with the use of rosiglitazone versus pioglitazone in several settings in observational studies (meta-analyses).

Monotherapy studies												
CV events	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	**I^2**
MI	0	–	–	–	3	1.23 (0.75, 2.01)	0.413	76.20%	3	1.23 (0.75, 2.01)	0.413	76.20%
Stroke	0	–	–	–	1	1.33 (0.89, 1.98)	0.162	–	1	1.33 (0.89, 1.98)	0.162	–
CHF	0	–	–	–	1	0.84 (0.52, 1.35)	0.474	–	1	0.84 (0.52, 1.35)	0.474	–
CV death	0	–	–	–	1	0.93 (0.21, 4.12)	0.924	–	1	0.93 (0.21, 4.12)	0.924	–
All-cause death	0	–	–	–	2	1.06 (0.64, 1.74)	0.827	21.70%	2	1.06 (0.64, 1.74)	0.827	21.70%
Add-on studies												
CV events	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	**I^2**
MI	2	1.13 (0.77, 1.65)	0.537	4.40%	11	**1.11 (1.00, 1.22)**	0.042	55.10%	13	**1.11 (1.00, 1.22)**	0.034	48.50%
Stroke	0	–	–	–	5	**1.12 (1.00, 1.24)**	0.048	30.10%	5	**1.12 (1.00, 1.24)**	0.048	30.10%
CHF	0	–	–	–	8	**1.15 (1.04, 1.26)**	0.005	64.90%	8	**1.15 (1.04, 1.26)**	0.005	64.90%
CV death	0	–	–	–	0	–	–	–	0	–	–	–
All-cause death	0	–	–	–	6	**1.15 (1.09, 1.20)**	<0.001	0%	6	**1.15 (1.09, 1.20)**	<0.001	0%
Overall												
CV events	Case-control				Cohort				Overall			
	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	I^2	n	OR (95% CI)	*p-value*	**I^2**
MI	2	1.12 (0.78, 1.59)	0.538	4.38%	15	**1.12 (1.01, 1.25)**	0.027	60.29%	17	**1.12 (1.02, 1.30)**	0.019	55.93%
Stroke	0	–	–	–	6	**1.13 (1.02, 1.25)**	0.015	20.89%	6	**1.13 (1.02, 1.25)**	0.015	20.89%
CHF	0	–	–	–	9	**1.13 (1.09, 1.25)**	0.011	63.71%	9	**1.13 (1.09, 1.25)**	0.011	63.71%
CV death	0	–	–	–	1	0.93 (0.21, 4.12)	0.924	0%	1	0.93 (0.21, 4.12)	0.924	0%
All-cause death	0	–	–	–	8	**1.15 (1.09, 1.20)**	<0.001	0%	8	**1.15 (1.09, 1.20)**	<0.001	0%

CHF, congestive heart failure; CV, cardiovascular; MI, myocardial infarction; n, number of studies (one publication could provide more than one estimate and each estimate was considered one study for the total n) OR, odds ratio.
Bold values are statistically significant (95% Confidence Interval).

II.7.2. SUPPLEMENTAL DATA II.6 – LIST OF OBSERVATIONAL STUDIES INCLUDED IN THE META-ANALYSES

AZOULAY, L. ET AL., 2010. THIAZOLIDINEDIONES AND THE RISK OF INCIDENT STROKES IN PATIENTS WITH TYPE 2 DIABETES: A NESTED CASE-CONTROL STUDY. PHARMACOEPIDEMIOL DRUG SAF, 19(4), PP. 343-350.

BILIK, D., ET AL., 2010. THIAZOLIDINEDIONES, CARDIOVASCULAR DISEASE AND CARDIOVASCULAR MORTALITY: TRANSLATING RESEARCH INTO ACTION FOR DIABETES (TRIAD). PHARMACOEPIDEMIOL DRUG SAF, 19, PP. 715-721.

BREUNIG, I.M., ET AL., 2014. DEVELOPMENT OF HEART FAILURE IN MEDICAID PATIENTS WITH TYPE 2 DIABETES TREATED WITH PIOGLITAZONE, ROSIGLITAZONE, OR METFORMIN. J MANAG CARE PHARM, 20(9), PP. 895-903.

BROWNSTEIN, J.S., ET AL., 2010. RAPID IDENTIFICATION OF MYOCARDIAL INFARCTION RISK ASSOCIATED WITH DIABETES MEDICATIONS USING ELECTRONIC MEDICAL RECORDS. DIABETES CARE, 33(3), PP. 526-531.

CHOU, C.C., ET AL., 2011. INCIDENCE OF CARDIOVASCULAR EVENTS IN WHICH 2 THIAZOLIDINEDIONES ARE USED AS ADD-ON TREATMENTS FOR TYPE 2 DIABETES MELLITUS IN A TAIWANESE POPULATION. CLIN THER, 33(12), PP. 1904-1913.

DORE, D.D., ET AL., 2009. ASSOCIATION BETWEEN EXTENT OF THIAZOLIDINEDIONE EXPOSURE AND RISK OF ACUTE MYOCARDIAL INFARCTION. PHARMACOTHERAPY, 29(7), PP. 775-783.

DORMUTH, C.R., ET AL., 2009. ROSIGLITAZONE AND MYOCARDIAL INFARCTION IN PATIENTS PREVIOUSLY PRESCRIBED METFORMIN. PLOS ONE, 4(6), PP. E6080.

GALLAGHER, A.M., ET AL., 2011. RISK OF DEATH AND CARDIOVASCULAR OUTCOMES WITH THIAZOLIDINEDIONES: A STUDY WITH THE GENERAL PRACTICE RESEARCH DATABASE AND SECONDARY CARE DATA. PLOS ONE, 6(12), PP. E28157.

GERRITS, C.M., ET AL., 2007. A COMPARISON OF PIOGLITAZONE AND ROSIGLITAZONE FOR HOSPITALIZATION FOR ACUTE MYOCARDIAL INFARCTION IN TYPE 2 DIABETES. PHARMACOEPIDEMIOL DRUG SAF, 16(10), PP. 1065-1071.

GRAHAM, D.J., ET AL., 2010. RISK OF ACUTE MYOCARDIAL INFARCTION, STROKE, HEART FAILURE, AND DEATH IN ELDERLY MEDICARE PATIENTS TREATED WITH ROSIGLITAZONE OR PIOGLITAZONE. JAMA, 304(4), PP. 411-418.

HABIB, Z.A., ET AL., 2009. RELATIONSHIP BETWEEN THIAZOLIDINEDIONE USE AND CARDIOVASCULAR OUTCOMES AND ALL-CAUSE MORTALITY AMONG PATIENTS WITH DIABETES: A TIME-UPDATED PROPENSITY ANALYSIS. PHARMACOEPIDEMIOL DRUG SAF, 18(6), PP. 437-447.

HSIAO, F.Y., ET AL., 2009. THIAZOLIDINEDIONES AND CARDIOVASCULAR EVENTS IN PATIENTS WITH TYPE 2 DIABETES MELLITUS: A RETROSPECTIVE COHORT STUDY OF OVER 473,000 PATIENTS USING THE NATIONAL HEALTH INSURANCE DATABASE IN TAIWAN. DRUG SAF, 32(8), PP. 675-690.

JUURLINK, D.N., ET AL., 2009. ADVERSE CARDIOVASCULAR EVENTS DURING TREATMENT WITH PIOGLITAZONE AND ROSIGLITAZONE: POPULATION BASED COHORT STUDY. BMJ, 18 (339), PP. B2942.

KARTER, A.J., ET AL., 2005. PIOGLITAZONE INITIATION AND SUBSEQUENT HOSPITALIZATION FOR CONGESTIVE HEART FAILURE. DIABET MED, 22(8), PP. 986-993.

KORO, C.E.; FU, Q.; & STENDER, M.; 2008. AN ASSESSMENT OF THE EFFECT OF THIAZOLIDINEDIONE EXPOSURE ON THE RISK OF MYOCARDIAL INFARCTION IN TYPE 2 DIABETIC PATIENTS. PHARMACOEPIDEMIOL DRUG SAF, 17(10), PP. 989-986.

LIPSCOMBE, L.L., ET AL., 2007. THIAZOLIDINEDIONES AND CARDIOVASCULAR OUTCOMES IN OLDER PATIENTS WITH DIABETES. JAMA, 298(22), PP. 2634-2643.

LOEBSTEIN, R., ET AL., 2011. DATABASE EVALUATION OF LONG-TERM ROSIGLITAZONE TREATMENT ON CARDIOVASCULAR OUTCOMES IN PATIENTS WITH TYPE 2 DIABETES. J CLIN PHARMACOL, 51, PP. 173-180.

LU, C.J., ET AL., 2013. RISK OF STROKE WITH THIAZOLIDINEDIONES: A TEN-YEAR NATIONWIDE POPULATION-BASED COHORT STUDY. CEREBROVASC DIS, 36(2), PP. 145-51.

MARGOLIS, D.J.; HOFFSTAD, O.; & STROM, B.L.; 2008. ASSOCIATION BETWEEN SERIOUS ISCHEMIC CARDIAC OUTCOMES AND MEDICATIONS USED TO TREAT DIABETES. PHARMACOEPIDEMIOL DRUG SAF, 17(8), PP. 753-759.

MCAFEE, A.T., ET AL., 2007. CORONARY HEART DISEASE OUTCOMES IN PATIENTS RECEIVING ANTIDIABETIC AGENTS. PHARMACOEPIDEMIOL DRUG SAF, 16(7), PP. 711-725.

MORGAN C.L., ET AL., 2012. WHAT NEXT AFTER METFORMIN? A RETROSPECTIVE EVALUATION OF THE OUTCOME OF SECOND-LINE, GLUCOSE-LOWERING THERAPIES IN PEOPLE WITH TYPE 2 DIABETES. J CLIN ENDOCRINOL METAB, 97(12), PP. 4605-4612.

PANTALONE, K.M., ET AL., 2009. THE RISK OF DEVELOPING CORONARY ARTERY DISEASE OR CONGESTIVE HEART FAILURE, AND OVERALL MORTALITY, IN TYPE 2 DIABETIC PATIENTS RECEIVING ROSIGLITAZONE, PIOGLITAZONE, METFORMIN, OR SULFONYLUREAS: A RETROSPECTIVE ANALYSIS. ACTA DIABETOL, 46(2), PP. 145-154.

RAJAGOPALAN, R., ET AL., 2004. ASSOCIATION BETWEEN CONGESTIVE HEART FAILURE AND HOSPITALIZATION IN PATIENTS WITH TYPE 2 DIABETES MELLITUS RECEIVING TREATMENT WITH INSULIN OR PIOGLITAZONE: A RETROSPECTIVE DATA ANALYSIS. CLIN THER, 26(9), PP. 1400-1410.

RAMIREZ, S.P., ET AL., 2009. ROSIGLITAZONE IS ASSOCIATED WITH MORTALITY IN CHRONIC HEMODIALYSIS PATIENTS. J AM SOC NEPHROL, 20(5), PP. 1094-1101.

STOCKL, K.M., ET AL., 2009. RISK OF ACUTE MYOCARDIAL INFARCTION IN PATIENTS TREATED WITH THIAZOLIDINEDIONES OR OTHER ANTIDIABETIC MEDICATIONS. PHARMACOEPIDEMIOL DRUG SAF, 18(2), PP. 166-174.

TANNEN R, ET AL., 2013. A NEW "COMPARATIVE EFFECTIVENESS" ASSESSMENT STRATEGY USING THE THIN DATABASE: COMPARISON OF THE CARDIAC COMPLICATIONS OF PIOGLITAZONE AND ROSIGLITAZONE. PHARMACOEPIDEMIOL DRUG SAF, 22(1), PP. 86-97.

TZOULAKI, I., ET AL., 2009. RISK OF CARDIOVASCULAR DISEASE AND ALL CAUSE MORTALITY AMONG PATIENTS WITH TYPE 2 DIABETES PRESCRIBED ORAL ANTIDIABETES DRUGS: RETROSPECTIVE COHORT STUDY USING UK GENERAL PRACTICE RESEARCH DATABASE. BMJ, 339, PP. B4731.

VALLARINO C, ET AL., 2013. COMPARING PIOGLITAZONE TO INSULIN WITH RESPECT TO CANCER, CARDIOVASCULAR AND BONE FRACTURE ENDPOINTS, USING PROPENSITY SCORE WEIGHTS. CLIN DRUG INVESTIG, 33(9), PP. 621-631.

VANASSE, A., ET AL., 2009. STROKE AND CARDIOVASCULAR MORBIDITY AND MORTALITY ASSOCIATED WITH ROSIGLITAZONE USE IN ELDERLY DIABETIC PATIENTS. DIAB VASC DIS RES, 6(2), PP. 87-93.

WALKER, A.M.; KORO, C.E.; & LANDON, J.; 2008. CORONARY HEART DISEASE OUTCOMES IN PATIENTS RECEIVING ANTIDIABETIC AGENTS IN THE PHARMETRICS DATABASE 2000-2007. PHARMACOEPIDEMIOL DRUG SAF, 17(8), PP. 760-768.

WHEELER, S., ET AL., 2013. MORTALITY AMONG VETERANS WITH TYPE 2 DIABETES INITIATING METFORMIN, SULFONYLUREA OR ROSIGLITAZONE MONOTHERAPY. DIABETOLOGIA, 56(9), PP. 1934-43.

WERTZ, D.A., ET AL., 2010. RISK OF CARDIOVASCULAR EVENTS AND ALL-CAUSE MORTALITY IN PATIENTS TREATED WITH THIAZOLIDINEDIONES IN A MANAGED-CARE POPULATION. CIRC CARDIOVASC QUAL OUTCOMES, 3(5), PP. 538-545.

WINKELMAYER, W.C., ET AL., 2008. COMPARISON OF CARDIOVASCULAR OUTCOMES IN ELDERLY PATIENTS WITH DIABETES WHO INITIATED ROSIGLITAZONE VS PIOGLITAZONE THERAPY. ARCH INTERN MED, 168(21), PP. 2368-2375.

YANG, J., ET AL., 2014. A COMPARISON OF ALL-CAUSE MORTALITY WITH PIOGLITAZONE AND INSULIN IN TYPE 2 DIABETES: AN EXPANDED ANALYSIS FROM A RETROSPECTIVE COHORT STUDY. CURR MED RES OPIN, 30(11), 2223-2231.

ZIYADEH, N., ET AL., 2009. THE THIAZOLIDINEDIONES ROSIGLITAZONE AND PIOGLITAZONE AND THE RISK OF CORONARY HEART DISEASE: A RETROSPECTIVE COHORT STUDY USING A US HEALTH INSURANCE DATABASE. CLIN THER, 31(11), PP. 2665-2677.

CHAPTER III – TESTING THE USEFULNESS OF THE NUMBER NEEDED TO TREAT TO BE HARMED (NNTH) IN BENEFIT-RISK EVALUATIONS: CASE STUDY WITH MEDICINES WITHDRAWN FROM THE EUROPEAN MARKET DUE TO SAFETY REASONS

III. TESTING THE USEFULNESS OF THE NUMBER NEEDED TO TREAT TO BE HARMED (NNTH) IN BENEFIT-RISK EVALUATIONS: CASE STUDY WITH MEDICINES WITHDRAWN FROM THE EUROPEAN MARKET DUE TO SAFETY REASONS

III.1. ABSTRACT

The objective of the study was to explore the usefulness of number needed to treat to be harmed (NNTH), in benefit-risk assessments, by studying the agreement between NNTH values and withdrawals of medicines from European market due to safety reasons. Medicines with data from longitudinal studies were included. Studies were identified from European Medicines Agency's Reports. Meta-analyses were performed to pool odds ratios (OR) with 95% confidence-intervals (CI). Published control event rates were applied to ORs to calculate NNTHs (95%CI) for selected adverse events. NNTH (95%CI) decreased from pre- to post-marketing for the eight medicines included: peripheral neuropathy (∞ vs. 12 [non-significant; NS] with almitrine; heart valve disease with benfluorex (∞ vs. NNTH ranging from 7 [4–13] to 7 [5–9]); myopathy (-4096 [NS] vs. 797 [421–1690]), new-onset diabetes (113 [NS] vs. 390 [425–778]), bleeding (∞ vs. 517 [317–1153]), and infection (∞ vs. 253 [164–463]) with niacin-laropiprant; psychiatric disorders (12 [7–34] vs. 9 [5–24]) with rimonabant; myocardial infarction (MI) [-1305 vs. 270 [89–4362]) with rofecoxib; MI (− 510 vs. NNTH ranging from 152 [55–4003] to 568 [344–1350]) with rosiglitazone; cardiovascular events (∞ vs. 245 [129–1318]) with sibutramine; and liver injury (∞ vs. 5957 [NS]) with ximelagatran. In conclusion, NNTH have potential of use as a supportive tool in benefit-risk re-evaluations of medicines and may help regulators to making decisions on drug safety.

III.2. INTRODUCTION

The assessment of benefit–risk (BR) ratios is a complex process based on the evaluation of the best evidence available about the efficacy and safety of medicines (Vandenbroucke & Psaty 2008), (Hammad et al. 2013). The evaluation of efficacy is often reduced to a one-dimension variable, which is well defined in randomized controlled trials (RCTs) specifically designed to detect differences between interventions on that parameter (Stanley 2007), (Singal, Higgins & Waljee 2014). However, the assessment of safety is more challenging, since it may comprise several harmful effects that can arise from numerous sources of evidence (Curtin & Schulz 2011), (Singh & Loke 2012), (Zorzela et al. 2014), (Alves, Batel-Marques & Macedo 2012), (Alves, Macedo & Batel-Marques 2013), (Alves, Batel-Marques & Macedo 2014). According to the Cochrane Adverse Effects Methods Group, it is not possible to make recommendations regarding the types of studies that must be considered in a systematic review of adverse drug reactions (Loke et al. 2007). The assessment of safety should comprehend a broad review of the evidence without restricting the analysis to certain study designs (Golder, Loke & Bland 2013). While there are relatively well-established methodologies for assessing efficacy, further investigation is particularly needed with regard to the development of more appropriate methodologies for evaluating safety.

The BR ratio assessment is essentially a subjective qualitative weighing process of the available evidence. Thus, variations in clinical and scientific judgments among experts can lead to different conclusions regarding the balance of benefits and risks (FDA 2013a). Thus, regulatory authorities may reach different decisions based on the same data, as it was the case of rosiglitazone (EMA 2010b), (FDA 2011), (Mendes, Alves & Batel-Marques 2015). In this context, regulatory authorities and pharmaceutical companies have initiated projects aimed to build frameworks that could serve as standardized structured models for BR assessment to achieve transparency in decision-making (FDA 2013a), (EMA 2009a). The PROTECT Project, which is coordinated by the European Medicines Agency (EMA), is an example aiming at developing and testing tools and processes for balancing benefits and risks, which could further be used as an aid to make informed and science-based regulatory decisions (EMA 2009a).

The authors of a systematic review about methodologies for BR assessment recommended 13 methodologies for further examinations, including the number needed to treat (NNT) (Mt-Isa et al. 2014). NNT is a measure of effect size that is defined as the

number of patients who need to be treated with one therapy versus another in order to encounter an additional outcome of interest over a defined period of time (Laupacis, Sackett & Roberts 1988), (Cook & Sackett 1995). NNT can be calculated for both beneficial and harmful events. In order to indicate direction of effect, two preferred notations are used, namely 'number needed to treat for an additional beneficial outcome' (NNTB) and 'number needed to treat for an additional harmful outcome' (NNTH). Thus, NNTB and NNTH can be used to assess benefits and risks of drugs, respectively.

While the use of metric indices has proven to have value in daily clinical practice, namely at assisting physicians in selecting therapeutic interventions (Straus et al. 2011), (Citrome & Ketter 2013), its usefulness in BR assessments involving drug regulation is yet to be established (Mendes, Alves & Batel-Marques 2015), (Mt-Isa et al. 2014). The use of the NNTH in BR ratio assessments can be important as it represents an absolute measure of effect (Citrome 2010). Relative measures of potential harm, such as relative risk (RR), odds ratio (OR), and hazard ratio (HR), are more commonly seen in the scientific literature (Citrome 2010). However, they do not reflect the risk of the outcome of interest without therapy (baseline risk) and, therefore, it is not possible to discriminate huge from small treatment effects (Straus et al. 2011). For example, if the incidence over a year of treatment for a serious adverse event is rare (0.03%) or frequent (30%) in group A and similarly rare (0.01%) or frequent (10%) in group B, the RR will always be 3.0. Although the difference is statistically significant in both scenarios, the clinical relevance is totally different. The interpretation of its relevance may implicate or justify the use of absolute measures, such as the NNTH. Based on such example, one would have one additional serious adverse event in each 5000 or 5 patients treated over a year, depending on the scenario in analysis. This example illustrates the potential usefulness of metric indices for making decisions on medicines evaluation.

The main purpose of this study was to explore the usefulness of the NNTH by studying the agreement between NNTH values and decisions of withdrawing medicines from the market due to unacceptable safety hazards. Theoretically, NNTHs should be lower in the post-marketing period for those medicines. To test this hypothesis, information from pre- and post-marketing studies was collected in order to carry out comparisons between the two points of time.

III.3. METHODS

III.3.1. IDENTIFICATION OF MEDICINES

Medicines suspended or withdrawn from the European market due to safety reasons between 2001 and 2015 were considered for inclusion in this study. The website of the EMA was searched in order to identify the medicines, namely the 'News, press release and public statement archive' and the 'Referrals' archive. Medicines were included in the study irrespectively of the withdrawal request has been made by the marketing authorization (MA) holder or the EMA. Medicines suspended during the study period that were reintroduced later in the market were not included.

Medicines withdrawn from the market based on data derived from longitudinal controlled studies, i.e. meta-analyses, RCTs and cohort studies, were selected for further quantitative analyses since they allow for estimating rates of events and consequently the calculation of NNTH values.

III.3.2. QUANTITATIVE ANALYSES

The underlying assumption was that the NNTHs should decrease over time. The BR would be positive when market authorizations were granted, but the risks would outweigh the benefits by the time of the withdrawals from the market. Thus, NNTB was assumed to be constant over time.

III.3.2.1. Time intervals

Data analysis was carried out for two points of time for each drug: (1) pre-marketing and (2) post-marketing. The pre-marketing period comprehended data obtained from RCTs conducted before the granting of the MA. The post-marketing period included data obtained from studies conducted between the date of MA and the date of market withdrawal.

III.3.2.2. Data sources

III.3.2.2.1. Pre-marketing

Searches were primarily conducted on the website of the EMA to identify RCTs in the pre-marketing European Public Assessment Reports (EPARs) issued before the approval of medicines, later withdrawn from the market. For drugs approved before the creation of the EMA or by mutual recognition procedure (MRP), marketing authorization holders (MAHs) were contacted to provide data on RCTs that supported the introduction of medicines in the EU market.

III.3.2.2.2. Post-marketing

Documents prepared by the EMA following post-marketing BR ratios re-assessments of medicines, namely 'press releases', 'questions and answers,' and 'scientific conclusions,' were used to identify and extract data from the studies that supported withdrawal decisions. Data were obtained from full-papers published in the literature if not available in the documents published by the Agency.

III.3.2.3. Data extraction and analyses: numbers needed to harm

Study design, study duration, interventions, comparators, number of randomized patients, and number of adverse events of interest (i.e. those that supported withdrawal decisions) were extracted from (1) the pre-marketing RCTs reviewed by Regulatory Authorities before granting a MA and (2) from the post-marketing studies used to support the decision of withdrawing a drug from the market. Data were pooled from all studies cited in documents issued by Regulatory Authorities. Data on all doses studied during the clinical development of the withdrawn drugs were considered. Data on all comparators (including placebo and/or active comparators) used during the clinical development were also considered. Only dichotomous events were considered (number of events in group A versus number of events in group B).

III.3.2.4. Data analysis and NNTH

Usually, NNT is calculated by taking the reciprocal of the absolute risk difference between two groups when appraising dichotomous data from a single RCT (Straus et al. 2011). The traditional way of calculating NNT should be applied only when analysing data from single studies or few studies with identical follow-up times. However, since data could be obtained from multiple studies, meta-analyses were performed to determine pooled evidence from RCTs and observational studies whenever applicable. As absolute risk differences are most likely to vary across different baseline event rates, they may be less appropriate for calculating NNTs from meta-analyses (Deeks, Higgins & Altman 2011). Relative effects tend to be more stable across risk groups than does absolute differences (Deeks, Higgins & Altman 2011). Thus, in the present study, NNTH values (and 95% CIs) were estimated for each adverse event of interest by applying pooled ORs (and the limits of its corresponding 95% CI) from meta-analyses (or individual studies when applicable) to annual control event rates (CERs) (Straus et al. 2011).

Meta-analyses were conducted using a random-effects model in order to pool the OR with their 95% CIs when there were at least two studies with the same design reporting on the adverse events of interest (DerSimonian & Laird 1986). This model was chosen as it is more conservative than a fixed-effect model in the presence of between-studies heterogeneity. Between-studies heterogeneity was assessed using the I^2 measure of inconsistency (Higgins et al. 2003). If there was only one study available for the adverse event of interest, the risk estimate (RR, OR, or HR) provided in that study was used to calculate NNTH values.

Because the adverse events assessed in the present study can be considered as rare, similarity was assumed between RR, OR, and HR (Mendes, Alves & Batel-Marques 2015). The most adjusted estimates were used for studies presenting more than one risk estimate. Annual CERs for adverse events were obtained from the published literature. The formula used to calculate NNTH values from OR results was the following: NNTH = 1 + [(CER) × (OR− 1)]/[(1− CER) × (CER) × (OR− 1)] (Straus et al. 2011). The NNTH was rounded up to next whole number. In case of the 95% CI for a NNTH estimate contain infinity (i.e. when the 95% CI for the OR contain zero), NNTH is not statistically significant at the *p* threshold of <0.05. In such situation, one of the confidence limits indicates harm and the other will indicate benefit. The confidence limits should therefore be labelled as NNTH and NNTB to express the direction of effect. Thus, the scale for NNT goes from NNTH = 1 to NNTB = 1

via infinity (Altman 1998). Further, for adverse events that were not detected or reported in pre-marketing period studies, NNTH was considered as 'infinite' by default. Analyses were performed using Stata version 12 (StataCorp, College Station, Texas).

III.4. RESULTS

The search strategy identified 27 medicines. Figure III. 1 presents the flowchart of the study. Nineteen medicines were excluded from further analyses because the available data were not enough to calculate NNTHs (Figure III. 1). Eight medicines were included for further analyses: almitrine, benfluorex, nicotinic acid/laropiprant, rimonabant, rofecoxib, rosiglitazone, sibutramine, and ximelagatran. The therapeutic indications, the safety concerns, and the list of evidence that supported the withdrawal of medicines from the market are presented in Table III. 1.

Table III. 2 presents medicines withdrawn from market, adverse events of interest, type of studies that were used to extract data, number of patients and events, ORs, CERs, and NNTHs. Further details on the studies used to extract data for all medicines are provided in Supplemental Table III. 1. Pre-marketing studies could not be identified for almitrine and benfluorex.

III.4.1. ALMITRINE

The NNTH calculated for almitrine versus placebo on peripheral neuropathy was estimated at 12 (95% CI NNTH 4 to ∞ to NNTB 819) based on data from post-marketing RCTs.

III.4.2. BENFLUOREX

According to post-marketing data, benfluorex was associated with an increased risk for heart valve diseases in a RCT versus pioglitazone (NNTH 7, 95% CI 4–13) and in a cohort study versus non-exposure (NNTH 7, 95% CI 5–9).

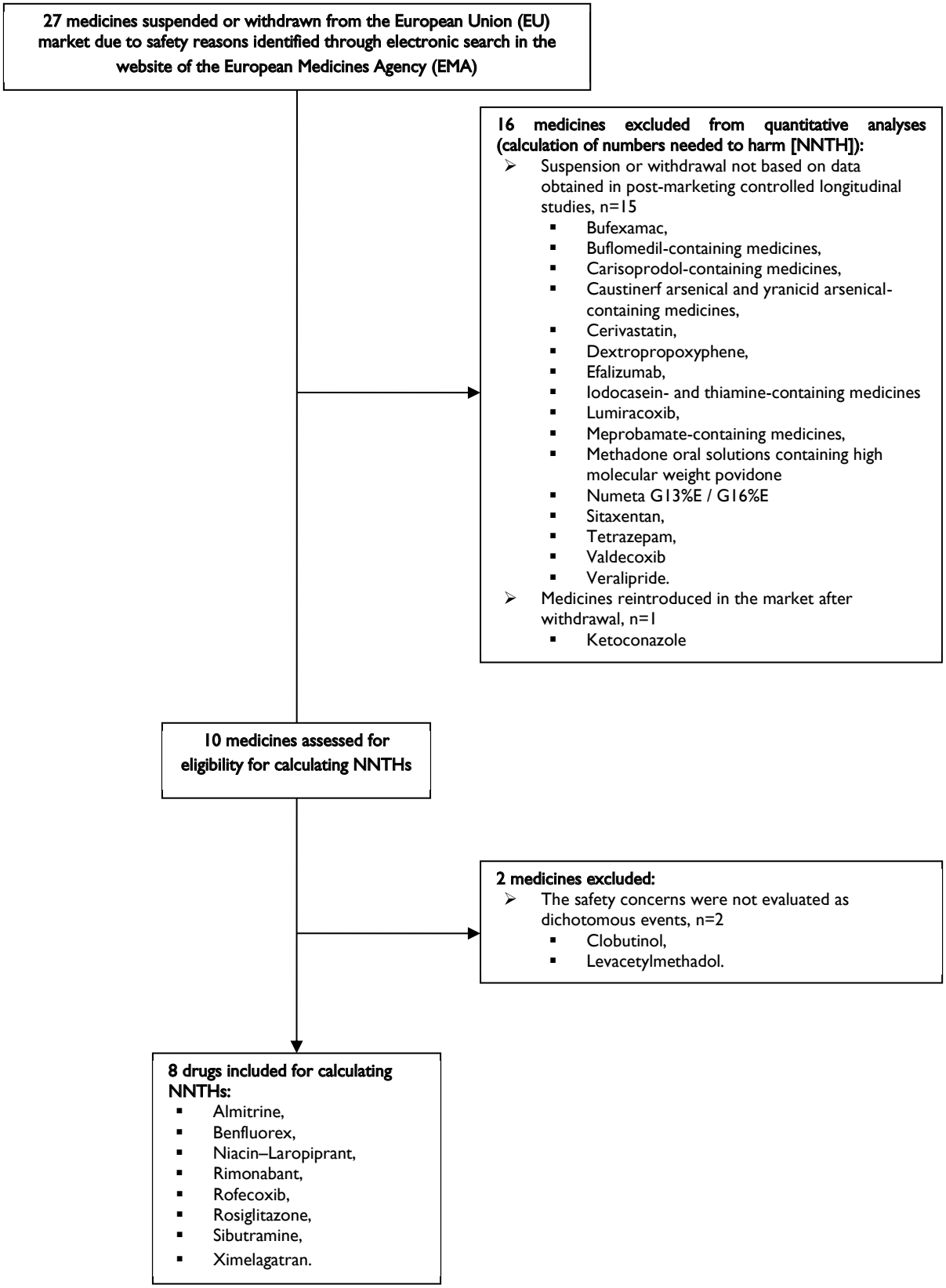


Figure III. 1 – Flowchart of the study.

III.4.3. NICOTINIC ACID/LAROPIPRANT

Based on data pooled from pre-marketing RCTs, NNTHs were estimated for nicotinic acid/laropiprant (±simvastatin) versus comparators (placebo, nicotinic acid alone, or simvastatin) on myopathy (NNTB 4096, 95% CI NNTB 2768 to ∞ to NNTH 380) and on new-onset diabetes (NNTH 113, 95% CI NNTH 23 to ∞ to NNTB 353). Pre-marketing data were not available on both serious bleeding and serious infection.

According to the post-marketing HPS2-THRIVE Trial, the use of nicotinic acid/laropiprant versus placebo resulted in a statistically significant increased risk of myopathy (NNTH 797, 95% CI 451–1690), new-onset diabetes (NNTH 390, 95% CI 245–778), serious bleeding (NNTH 517, 95% CI 317–1153), and serious infection (NNTH 253, 95% CI 164–463).

III.4.4. RIMONABANT

Rimonabant was associated with a statistically significant increased risk of psychiatric disorders (leading to treatment discontinuation) versus placebo in pre-marketing RCTs (NNTH 12, 95% CI 7–34). As compared with pre-marketing studies, a slight decrease in the NNTH value was noted in the post-marketing STRADIVARIUS Trial (NNTH 9, 95% CI 5–24).

III.4.5. ROFECOXIB

According to data from pre-marketing RCTs, the effect of rofecoxib was not different from its comparators (placebo or NSAIDs) with regards to the risk of myocardial infarction (MI) (NNTB 1305, 95% CI NNTB 478 to ∞ to NNTH 293).

The results of the post-marketing APPROVe Trial revealed a statistically significant increased risk of MI with rofecoxib versus placebo (NNTH 270, 95% CI 89–4362).

III.4.6. ROSIGLITAZONE

Data from pre-marketing RCTs cited in the EPAR were used to estimate the NNTH for rosiglitazone versus comparators on MI (NNTB 510, 95% CI NNTB 173 to ∞ to NNTH 134).

The post-marketing re-assessment of the BR ratio of rosiglitazone carried out by the EMA was based on information from several sources of evidence. A statistically significant increased risk of MI with rosiglitazone versus non-thiazolidinedione comparators were found in two meta-analyses of RCTs, one from the FDA (NNTH 152, 95% CI 55–4003), and another from Nissen and Wolski (NNTH 430, 95% CI 192–6004). The post-marketing RECORD Trial did not show an increased risk of MI with rosiglitazone versus comparators. The meta-analysis of the cohort studies cited in the systematic review analysed by the EMA indicated an increased risk of MI with rosiglitazone versus pioglitazone, with an overall statistically significant NNTH estimated at 568 (95% CI 344–1350). The EMA also highlighted in the assessment report the results obtained in the cohort study published by Graham et al., which indicated an increased risk on a composite cardiovascular outcome (MI, stroke, heart failure [HF], or all-cause mortality) with rosiglitazone versus pioglitazone (NNTH 60, 95% CI 48–79).

III.4.7. SIBUTRAMINE

Serious cardiovascular outcomes (MI, stroke, or cardiovascular mortality) were not reported during pre-marketing RCTs. In the post-marketing SCOUT Trial, sibutramine was associated with an increased risk versus placebo on the primary outcome event (nonfatal MI, nonfatal stroke, resuscitation after cardiac arrest, and cardiovascular mortality), with a resulting NNTH estimated at 248 (95% CI 129–1318).

III.4.8. XIMELAGATRAN

No cases of DILI were identified in the pre-marketing RCTs supplied by the MAH of ximelagatran. The post-marketing EXTEND Trial reported one case of DILI in the ximelagatran group and none in the enoxaparin group. The NNTH was estimated at 5957 (NNTH 105 to ∞ to NNTB 49486).

Table III. 1 – Medicines included in the study that were withdrawn from the EU market due to safety reasons, between 2001 and 2015.

Medicine	Therapeutic indication	Safety issue	Year of first marketing in Europe	Year of withdrawal	Evidence supporting withdrawal decision
Almitrine	Chronic respiratory failure, which is associated with hypoxaemia (e.g., COPD)	Significant weight loss and peripheral neuropathy (which can be long-lasting and possibly irreversible)	1982	2013	Case reports Clinical trials
Benfluorex	Add-on treatment in patients with diabetes who are overweight (in combination with an appropriate diet)	Heart valve disease and pulmonary arterial hypertension	1974	2009	Case reports Case-series Case-control studies Cohort studies Clinical trials
Nicotinic acid / laropiprant	Dyslipidaemia	Bleeding (intracranial and gastro-intestinal), myopathy, infections and new-onset diabetes	2008	2013	Clinical trials
Rimonabant	It is used together with diet and exercise to reduce weight in adult patients who are obese or overweight and also have other risk factors, such as type 2 diabetes or dyslipidaemia	Psychiatric disorders, particularly depression	2006	2009	Case reports Clinical trials
Rofecoxib	Symptomatic relief of rheumatoid arthritis, osteoarthritis, acute pain and pain due to primary dysmenorrhoea	Thrombotic events	1999	2004	Clinical trials
Rosiglitazone	Type 2 diabetes mellitus	Cardiovascular events, particularly myocardial infarction	2000	2010	Case reports Case-control studies Cohort studies Clinical trials Systematic review of observational studies Meta-analysis of clinical trials Others
Sibutramine	Weight-loss in obese patients and in overweight patients who also have other risk factors such as type-2 diabetes or dyslipidaemia, together with diet and exercise	Cardiovascular events, such as heart attack, stroke and cardiac arrest	1999	2010	Case reports Clinical trials
Ximelagatran	Prevention of stroke and other thromboembolic complications associated with atrial fibrillation	Liver injury	2003	2006	Clinical trials

Table III. 2 – Withdrawn drugs, adverse events of interest, odds ratio (OR), annual control event rates (CER) and numbers needed to treat to be harmed (NNTH) in pre- and post-marketing periods.

Drug / Period	Study Design	Adverse Event	Withdrawn Drug		Control		Odds Ratio (OR)	CER	NNTH (95% CI)
			Pts, N=	Ev, N=	Pts, N=	Ev, N=			
Almitrine		Peripheral neuropathy							
Pre-	NA		NA	NA	NA	NA	NA	6.90%[a]	NNTH = ∞
Post-	RCT[b]		559	71	533	19	2.49 (0.98 - 6.29)		11.6 (NNTH 4.0 to ∞ to NNTB 818.6)
Benfluorex		Heart valve disease							
Pre-	NA		NA	NA	NA	NA	NA	10.82%[c]	NNTH = ∞
Post-	RCT[d]		310	82	305	33	2.97 (1.91 - 4.63)		6.4 (4.0 - 12.5)
	Cohort[e]		43044	65	1006129	532	3.10 (2.40 - 4.00)		6.1 (4.6 - 8.5)
Nicotinic acid / laropiprant									
Pre-	RCT[f]	Myopathy	2327	0	2131	1	0.33 (0.014 - 8.21)	0.04%[h]	NNTB 4095.1 (NNTB 2765.7 to ∞ to NNTH 379.5)
Post-	RCT[g]		12838	75	12835	17	4.43 (2.62 - 7.51)		796.1 (420.4 – 1689.1)
Pre-	RCT[f]	New-onset diabetes	2327	9	2131	4	2.11 (0.65 - 6.87)	0.81%[h]	113.0 (NNTH 22.2 to ∞ to NNTB 352.1)
Post-	RCT[g]		12838	494	12835	376	1.32 (1.16 – 1.51)		389.5 (244.7 – 777.9)
Pre-	RCT[f]	Serious Bleeding	2327	NA	2131	NA	NA	0.51%[h]	NNTH = ∞
Post-	RCT[g]		12838	326	12835	238	1.38 (1.18 – 1.62)		516.2 (316.8 – 1152.7)
Pre-	RCT[f]	Serious Infection	2327	NA	2131	NA	NA	1.84%[h]	NNTH = ∞
Post-	RCT[g]		12838	1031	12835	853	1.22 (1.12 – 1.34)		252.7 (163.9 – 462.4)
Rimonabant		Psychiatric disorders[i]							
Pre-	RCT[k]		2503	157	1602	41	2.48 (1.49 - 4.12)	6.60%[j]	12.0 (6.3 - 33.9)
Post-	RCT[l]		422	40	416	13	3.25 (1.71 - 6.16)		8.3 (4.2 - 23.9)
Rofecoxib		Myocardial infarction							
Pre-	RCT[n]		2449	7	1558	5	0.72 (0.23 - 2.27)	0.27%[m]	NNTB 1304.2 (NNTB 477.9 to ∞ to NNTH 292.6)
Post-	RCT[o]		1287	21	1299	9	2.38 (1.09 - 5.21)		270.0 (89.0 - 4361.8)
Rosiglitazone		Myocardial infarction							
Pre-	RCT[q]		3614	12	1458	5	0.77 (0.31 - 1.91)	0.84%[p]	NNTB 509.9 (NNTB 172.2 to ∞ to NNTH 133.5)
Post-	FDA Meta-Analysis of RCTs		10039	45	6956	20	1.80 (1.03 - 3.25)		151.1 (54.4 – 4002.9)

Drug / Period	Study Design	Adverse Event	Withdrawn Drug		Control		Odds Ratio (OR)	CER	NNTH (95% CI)
			Pts, N=	Ev, N=	Pts, N=	Ev, N=			
Post-	Nissen and Wolski's Meta-Analysis of RCTs		17258	159	14449	136	1.28 (1.02 - 1.63)		429.8 (191.6 - 6003.8)
Post-	RCT[r]		2220	64	2227	56	1.14 (0.80 - 1.63)		858.6 (NNTH 191.6 to ∞ to NNTB 599.3)
Post-	Systematic Review of Cohort Studies[s]		187887	1788	168957	1325	1.21 (1.09 – 1.35)		567.3 (344.0 – 1350.0)
Post-	Cohort[t]		677593	2593	159978	5386	1.18 (1.12 – 1.23)		60 (48. 79)*
Sibutramine		CV events[u]							
Pre-	RCT[w]		1297	0	742	0	NA	2.60%[v]	NNTH = ∞
Post-	RCT[x]		4906	561	4898	490	1.16 (1.03 - 1.31)		247.8 (128.4 – 1317.3)
Ximelagatran		DILI							
Pre-	RCT[z]		7130	0	5182	0	NA	0.0024%[y]	NNTH = ∞
Post-	RCT[aa]		557	1	601	0	8.00 (0.16 - 404.14)		5956.9 (NNTH 104.4 to ∞ to NNTB 49485.5)

CER, control event rate; CV, cardiovascular; DILI, drug-induced liver injury; Ev, events; NA, not available; NNTB, number needed to treat to benefit; NNTH, number needed to treat to be harmed; OR, odds ratio; Post-, post-marketing; Pre-, pre-marketing; Pts, patients; RCT, randomized controlled trial.

a) CER for "peripheral neuropathy" from Gorecka, et al., 2003 (placebo group).
b) Data from Voisin, et al., 1987; Bardsley, et al., 1991; Weitzenblum, et al., 1992; Bardsley, et al., 1992; Gorecka, et al., 2003, (EMA 2013).
c) CER for "emergent regurgitation" from REGULATE Trial (Derumeaux, et al., 2012) (pioglitazone group).
d) Adverse event defined as "emergent regurgitation"; Derumeaux, et al., 2012.
e) Adverse event defined as "Risk of hospitalization for cardiac valvular insufficiency (cardiac valvular insufficiency for any cause, mitral insufficiency, and aortic insufficiency; valvular replacement surgery for valvular insufficiency of any cause"; Weill, et al., 2010.
f) Data from pivotal studies P020-02 (Maccubbin, et al., 2008), P022-02 (Gleim, et al., 2009) and P054-00 (Maccubbin, et al., 2009).
g) Data from HPS2-THRIVE Trial (HPS2-THRIVE Collaborative Group. 2014).
h) CERs for myopathy, new-onset diabetes, serious bleeding, and serious infection were all obtained from HPS2-THRIVE (placebo group).
i) Adverse event defined as "psychiatric disorder leading to treatment discontinuation".
j) CER for "Major Depressive Episode" from Center for Behavioral Health Statistics and Quality, 2015.
k) Data from RIO EUROPE (Van Gaal, et al., 2005), RIO LIPIDS (Després, et al., 2005), RIO NORTH AMERICA (Pi-Sunyer, et al., 2006), RIO DIABETES (Scheen, et al., 2006).
l) Data from STRADIVARIUS Trial (Nissen, et al., 2008).
m) CER for "myocardial infarction" from APPROVE Trial (placebo group) (Bresalier, et al., 2005).
n) Data from Osteoarthritis Trials 029 (Ehrich, et al., 2001), 034 (Saag, et al., 2000), 058 (Truitt, et al., 2001), 044 (Laine, et al., 1999), 045 (Hawkey, et al., 2000), and 035 (Cannon, et al., 2000).
o) Data from APPROVe Trial (Bresalier, et al., 2005).
p) CER for "myocardial infarction" from the Look AHEAD Research Group trial (Look AHEAD Research Group. 2013) (placebo group).
q) Data from RCTs 49653/006, 49653/011, 49653/015, 49653/020, 49653/024, 49653/079, 49653/090, 49653/093, 49653/094, 49653/096, and 49653/098 (EMA 2006a), (GSK 2016).
r) Data from RECORD Trial (Home, et al., 2009).
s) Data from cohort studies included in the systematic review assessing the risk of myocardial infarction (Brwonstein, et al., 2010), (Ziyadeh, et al., 2009), (Juurlink, et al., 2009), (Winkelmayer, et al., 2008), (Walker, et al., 2008), (Gerrits, et al., 2007).
t) Data from Graham, et al., 2010 on a composite outcome of myocardial infarction, stroke, congestive heart failure or all-cause mortality. *The NNTH was not calculated, but rather extracted from the publication.
u) Composite of nonfatal myocardial infarction, nonfatal stroke, resuscitation after cardiac arrest, or cardiovascular death.
v) CER for the "composite outcome in (u)" from SCOUT trial (James, et al., 2010) (placebo group).
w) Data from RCTs SB1047 (Smith, et al., 2001), SB1048 (James, et al., 2000), SB1049 (Apfelbaum, et al., 1999), KD9618 (Wirth, et al., 2001), SB5078 (Kaukua, et al., 2004), SB6085 (McNulty, et al., 2003).
x) Data from SCOUT trial (James, et al., 2010).
y) CER for "drug-induced liver injury" from de Abajo, et al., 2004.
z) Data from RCTs METHRO I (Eriksson, et al., 2002), METHRO II (Eriksson, et al., 2002), METHRO III (Dahl, et al., 2005), (Eriksson, et al., 2003), (Eriksson, et al., 2004), EXPRESS (Eriksson, et al., 2003), EXULT A (Francis, et al., 2003), and EXULT B (Colwell, et al., 2005).
aa) Data from EXTEND Trial (Agnelli, et al., 2009).

III.5. DISCUSSION

According to the overall results, a decrease in the value of the NNTH from the pre- to the post-marketing period of time was seen for all medicines included in the study, although statistically significant values have not been obtained with almitrine, and ximelagatran in post-marketing studies.

Due to the lack of data, pre-marketing NNTH values were not estimated for almitrine, nor benfluorex (introduced in the market in 1982 and 1974, respectively). In such situations, the pre-marketing NNTH was considered as infinite by default. Thus, there is at least new evidence generated in post-marketing studies that indicate increased risks for the adverse events of interest in association with both medicines, therefore supporting the withdrawal decisions made by the EMA. The NNTH estimated for peripheral neuropathy resulting from almitrine use did not reach statistically significance, but the lower limit of the 95% CI of the OR was very close to the unity. Moreover, a statistically significant NNTH [6.5, 95% CI 3.9–13.0] would have been encountered if a fixed-effects model had been used to pool the OR (3.85, 95% CI 2.30–6.46). Nevertheless, the random effects model OR has been recommended as the best summary measure for clinicians who need to calculate patient's expected event rate-adjusted NNT (Furukawa, Guyatt & Griffith 2002).

Nicotinic acid/Laropiprant was associated with lower values of NNTH for all adverse events analysed (myopathy, new-onset diabetes, serious bleeding, and serious infections) in the post-marketing HPS2-THRIVE study, as compared with pre-marketing studies (HPS2-THRIVE Collaborative Group 2014), (EMA 2008a), (EMA 2013b). Despite the point-estimate NNTH for new-onset diabetes was lower in pre-marketing studies, statistical significance was seen only during post-marketing. Further although statistically significant values of post-marketing NNTHs have been found for serious bleeding and serious infection, the lack of pre-marketing data precluded the calculation of NNTHs during that period of time. Overall, the evidence generated during post-marketing, with resulting statistically significant NNTHs, is in line with the conclusions and the decision made by the EMA.

In the case of rimonabant, there was a slight decrease in the NNTH estimated for psychiatric adverse events leading to treatment discontinuation from pre-marketing studies (NNTH 12) to the post-marketing STRADIVARIUS study (NNTH 9) (EMA 2006b), (Nissen et al. 2008). By the time of its approval, the EMA recognized that the use of rimonabant could increase the risk of psychiatric events, especially depression, but they concluded that the BR balance was positive even in the light of such events (EMA 2006b). Given the

uncertainties, they requested for additional research to clarify that risk (EMA 2006b). In contrast, the FDA panel of experts unanimously rejected the approval of rimonabant alleging that the drug nearly doubled the rate of psychiatric adverse events in RCTs where individuals with history of depression were mostly excluded from study (Dooren & Whalen 2007). The post-marketing NNTH confirms the initial suspicion and strengthens the association between the use of rimonabant and the development psychiatric adverse events.

Rofecoxib was first authorized in the United Kingdom in 1999, and thereafter in EU member states through a MRP (EMA 2004). According to the results of this study, data from pre-marketing RCTs suggest that rofecoxib had a neutral effect with regards to MI, as compared to control. However, when the results from the post-marketing APPROVe Trial became available, the EMA decided to withdraw rofecoxib from the market due to an increased risk of thrombotic events versus placebo (Bresalier et al. 2005), (EMA 2004). The NNTH estimated based on data from the APPROVe Trial for MI supports this conclusion. Nevertheless, regulatory authorities were criticized because the accumulated evidence available in late 2000 should have been enough to support the withdrawal decision (Jüni et al. 2004). As an example, data from the VIGOR Trial indicated a statistically significant increased risk of MI with rofecoxib versus naproxen, with a resulting NNTH of 126 (35–1125) (Bombardier et al. 2000).

The case of rosiglitazone provides an example that there is room for improvement in the clarity and reproducibility of BR assessments in drug review (Mendes, Alves & Batel-Marques 2015). EMA decided to withdraw rosiglitazone due to an increased cardiovascular risk (EMA 2010b), (EMA 2010c), (Blind et al. 2011). However, this medicine is still on the market in other countries, namely the USA (FDA 2011). In 2000, when rosiglitazone received MA in Europe, fluid retention and a possible increased risk of congestive HF were noted with the drug and the manufacturer was requested by the EMA to undertake a post-marketing long-term cardiovascular morbidity/mortality study (RECORD). The debate around the cardiovascular safety of rosiglitazone was intensified in 2007 after the publication of a meta-analysis of RCTs indicating an increased risk of MI and a trend toward increased cardiovascular mortality (Nissen & Wolski 2007). Similar analyses performed by the FDA and the manufacturer provided consistent results (EMA 2010b), (FDA 2010a), (FDA 2013b). Taking into account that the RECORD Trial was still ongoing, experts from the EMA considered that uncertainties remained regarding the cardiovascular safety of rosiglitazone (Blind et al. 2011). Definite results from the RECORD Trial were available in 2009, but they were inconclusive with regard to cardiovascular events (Home et al. 2009). The NNTH

estimated based on data from the RECORD Trial indicates a trend toward an increased risk of MI with rosiglitazone (NNTH 857; not statistically significant). The BR ratio of rosiglitazone was reassessed again in 2010 after the publication of a meta-analysis of RCTs and a large cohort study. According to data from the two meta-analyses of RCTs cited in the EMA's assessment report, NNTH values (152 and 430 in the FDA's meta-analysis and Nissen and Wolski's meta-analysis, respectively) decreased comparing with those estimated in pre-marketing RCTs (EMA 2010b), (FDA 2010a), (Nissen & Wolski 2010). Further, the results of a systematic review of observational studies and of a single large cohort study from Graham et al. (on a composite end point of MI, stroke, HF, or death) also pointed out an increased cardiovascular risk with rosiglitazone, as compared with pioglitazone (EMA 2010b), (Graham et al. 2010). According to the EMA's conclusions, the meta-analysis performed by Nissen and Wolski in 2010 and the observational study conducted by Graham and colleagues have particularly weighed in the final decision (EMA 2010b). The results obtained in the present study across the several sources of evidence corroborate the conclusions achieved by the agency.

Sibutramine was associated with an increased risk (and a statistically significant NNTH) for a composite outcome of serious cardiovascular events (MI, stroke, resuscitation after cardiac arrest, and cardiovascular mortality) in the post-marketing SCOUT study (James et al. 2010). The cardiovascular profile of sibutramine was discussed before the granting of MA due to changes in blood pressure and heart rate noted in pre-marketing trials (EMA 2001). Although high blood pressure is a foremost risk factor for major cardiovascular events (Franklin & Wong 2013), such episodes were not reported in pre-marketing studies (EMA 2001). Thus, the NNTH values found in the post-marketing setting provide reassuring evidence about effect of sibutramine on cardiovascular outcomes.

The pre-marketing studies submitted by the MAH of ximelagatran to European regulators reported no cases of severe liver injury (Eriksson et al. 2002a), (Eriksson et al. 2002b), (Dahl et al. 2005), (Eriksson et al. 2003a), (Eriksson et al. 2003b), (Eriksson et al. 2004), (Colwell et al. 2005). Though, such cases were observed in other pre-marketing studies not included in the European preapproval dossier, namely SPORTIF III and SPORTIF V (FDA 2004). Ximelagatran was never approved by the FDA due to concerns of hepatotoxicity (Astrazeneca 2003), (Jeffrey 2004). Instances of transaminase elevation accompanied by elevated bilirubin have often predicted post-marketing serious liver injuries, including fatalities and patients requiring transplantation (FDA 2000). Ximelagatran ended to be withdrawn from the EU market because of a single patient that developed severe liver

injury after ximelagatran withdrawal within the post-marketing EXTEND study (Agnelli et al. 2009), (AstraZeneca 2006). The authors of the EXTEND study concluded that regular liver function monitoring was not enough to prevent cases of liver injury (Agnelli et al. 2009). The usefulness of the NNTH seems to be limited in cases like ximelagatran because of the unpredictability of events of this nature. Although a trend for an increased risk of serious liver injury was seen in post-marketing studies, the NNTH value did not reach statistical significance due to the rarity of such event.

One constraint of the NNT (NNTB and NNTH) methodology is the dependence on the CER of the disease. The estimates are sensitive to different patient profiles and applicable only to populations whose baseline risk is similar to the study populations (Laupacis, Sackett & Roberts 1988), (Baglin 2009). Thus, patients with differing severity of disease will probably have a different baseline risk. In such cases, different CERs lead to different NNT values (Deeks, Higgins & Altman 2011). Clinicians, other health professionals, and even regulators, should therefore evaluate if the results are applicable to their patients, by taking into account the characteristics of the population included in the study used as source of the CER. If not, another CER need to be used to calculate NNTs. Some authors have recommended to express relative risks (e.g. OR, RR) and a variety of NNTs across a range of different CERs (McQuay & Moore 1997), (Smeeth, Haines & Ebrahim 1999).

This study presents some limitations. The underlying assumption of this study was that the NNTB would be constant over time. However, analyses were not performed to confirm such hypothesis, as the study focused only on the assessment adverse events. Although a deterioration of the safety profile was noted for all medicines over time, the extension in which a possible concomitant deterioration of the efficacy profiles may have contributed to unbalance BR ratios was not examined. For example in the case of rimonabant, in addition to the conclusion that 'serious psychiatric disorders may be more common than in the clinical trials used in the initial assessment of the medicine,' 'new data show that in real life, patients tend to stop their treatment early,' and that 'the short-term treatment may not bring the benefits expected on the basis of the clinical trials' (EMA 2008b). Further, the definition of the adverse events may not be entirely comparable between different studies. For example, in the case of heart valve disease with benfluorex, the REGULATE Trial reported cases of emergent regurgitation, while the cohort study reported all cases of hospitalization due to cardiac valvular insufficiency for any cause, mitral insufficiency, and aortic insufficiency (Derumeaux et al. 2012), (Weill et al. 2010). This type of inconsistency in the definition of the adverse events may result in detection and/or

selection bias that could affect the NNTH calculation, as well as preclude straightforward comparisons between different studies. In addition, premarketing rates of events were calculated by pooling data from several studies that could have used different controls (active and placebos) compared to those used in post-marketing studies. The exposition to different controls may influence results on rates differences and consequently on NNTH estimates. Lastly, only references cited in public assessment reports were considered. However, regulatory authorities may have reviewed other studies than those cited in the assessment reports and MAHs possibly sponsored other studies that are not publicly available. Publication bias in industry sponsored trials is also particularly prominent in the reporting of adverse drug events (Hughes, Cohen & Jaggi 2014), (Doshi & Jefferson 2013), (Golder & Loke 2008), (Potthast et al. 2014).

According to the overall results of this study, a reduction of NNTH values was noted from the pre-marketing to the post-marketing assessment periods of time for medicines withdrawn from the European market due to safety reasons. Therefore, NNTH have the potential to be used as a supportive tool in BR ratio re-evaluations of marketed drugs and may have value in assisting regulatory authorities to making decisions on drug safety. Nevertheless, further research is needed using other case studies, namely for medicines that are currently in the market.

III.6. REFERENCES

AGNELLI, G. ET AL., 2009. SAFETY ASSESSMENT OF NEW ANTITHROMBOTIC AGENTS: LESSONS FROM THE EXTEND STUDY ON XIMELAGATRAN. THROMB RES, 123(3), PP. 488-497.

ALTMAN, D., 1998. CONFIDENCE INTERVALS FOR THE NUMBER NEEDED TO TREAT. BMJ, 317(7168), PP. 1309-1312.

ALVES, C., BATEL-MARQUES, F. & MACEDO, A., 2012. DATA SOURCES ON DRUG SAFETY EVALUATION: A REVIEW OF RECENT PUBLISHED META-ANALYSES. PHARMACOEPIDEMIOL DRUG SAF, 21(1), PP. 21-33.

ALVES, C., MACEDO, A. & BATEL-MARQUES, F., 2013. SOURCES OF INFORMATION USED BY REGULATORY AGENCIES ON THE GENERATION OF DRUG SAFETY ALERTS. EUR J CLIN PHARMACOL, 69(12), PP. 2083-2094.

ALVES, C., BATEL-MARQUES, F. & MACEDO, A., 2014. DRUG-SAFETY ALERTS ISSUED BY REGULATORY AUTHORITIES: USEFULNESS OF META-ANALYSIS IN PREDICTING RISKS EARLIER. EUR J CLIN PHARMACOL, 70(6), PP. 745-756.

APFELBAUM, M. ET AL., 1999. LONG-TERM MAINTENANCE OF WEIGHT LOSS AFTER A VERY-LOW-CALORIE DIET: A RANDOMIZED BLINDED TRIAL OF THE EFFICACY AND TOLERABILITY OF SIBUTRAMINE. AM J MED, 106(2), 179-184.

ARNAUD, F. ET AL., 1983. LONG TERM ALMITRINE BISMESYLATE TREATMENT IN PATIENTS WITH CHRONIC BRONCHITIS AND EMPHYSEMA: A MULTICENTRE DOUBLE-BLIND PLACEBO CONTROLLED STUDY. EUR J RESPIR DIS SUPPL, 126, PP. 323-336.

ASTRAZENECA, 2003. EXANTA (XIMELAGATRAN) RECEIVES FIRST APPROVAL [PRESS RELEASE: 23-12-2003]. [ONLINE] AVAILABLE AT: HTTP://195.69.152.75/MEDIA/PRESS-RELEASES/ARTICLE/20031223--EXANTA-XIMELAGATRAN-RECEIVES-FIRST-APPROVAL [ACCESSED 09 DECEMBER 2016].

ASTRAZENECA, 2006. LETTER: WITHDRAWAL OF XIMELAGATRAN 36 MG FILM-COATED TABLETS: EMEA/H/C/000702 [14-02-2006]. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/OTHER/2010/01/WC500069726.PDF [ACCESSED 03 MAY 2016].

BAGLIN, T., 2009. COMMUNICATING BENEFIT AND RISK. BR J HAEMATOL, 146(1), PP. 31-33.

BAKRAN, I. ET AL., 1990. DOUBLE-BLIND PLACEBO CONTROLLED CLINICAL TRIAL OF ALMITRINE BISMESYLATE IN PATIENTS WITH CHRONIC RESPIRATORY INSUFFICIENCY. EUR J CLIN PHARMACOL, 38(3), PP. 249-253.

BARDSLEY, P.A. ET AL., 1991. TWO YEARS TREATMENT WITH ALMITRINE BISMESYLATE IN PATIENTS WITH HYPOXIC CHRONIC OBSTRUCTIVE AIRWAYS DISEASE. EUR RESPIR J, 4(3), PP. 308-310.

BARDSLEY, P.A. ET AL., 1992. SEQUENTIAL TREATMENT WITH LOW DOSE ALMITRINE BISMESYLATE IN HYPOXAEMIC CHRONIC OBSTRUCTIVE AIRWAYS DISEASE. EUR RESPIR J, 5(9), PP. 1054-1061.

BLIND, E., DUNDER, K., DE GRAEFF, P. & ABADIE, E., 2011. ROSIGLITAZONE: A EUROPEAN REGULATORY PERSPECTIVE. DIABETOLOGIA, 54(2), PP. 213-218.

BOMBARDIER, C. ET AL., 2000. COMPARISON OF UPPER GASTROINTESTINAL TOXICITY OF ROFECOXIB AND NAPROXEN IN PATIENTS WITH RHEUMATOID ARTHRITIS. VIGOR STUDY GROUP. N ENGL J MED, 343(21), PP. 1520-8, 2 P FOLLOWING 1528.

BRESALIER R.S., ET AL., 2005. CARDIOVASCULAR EVENTS ASSOCIATED WITH ROFECOXIB IN A COLORECTAL ADENOMA CHEMOPREVENTION TRIAL. N ENGL J MED, 352(11), PP. 1092-1102.

BROWNSTEIN, J.S. ET AL., 2010. RAPID IDENTIFICATION OF MYOCARDIAL INFARCTION RISK ASSOCIATED WITH DIABETES MEDICATIONS USING ELECTRONIC MEDICAL RECORDS. DIABETES CARE, 33(3), PP. 526-531.

CANNON, G.W. ET AL., 2000. ROFECOXIB, A SPECIFIC INHIBITOR OF CYCLOOXYGENASE 2, WITH CLINICAL EFFICACY COMPARABLE WITH THAT OF DICLOFENAC SODIUM: RESULTS OF A ONE-YEAR, RANDOMIZED, CLINICAL TRIAL IN PATIENTS WITH OSTEOARTHRITIS OF THE KNEE AND HIP. ROFECOXIB PHASE III PROTOCOL 035 STUDY GROUP. ARTHRITIS RHEUM, 43(5), PP. 978-987.

CENTER FOR BEHAVIORAL HEALTH STATISTICS AND QUALITY, 2015. BEHAVIORAL HEALTH TRENDS IN THE UNITED STATES: RESULTS FROM THE 2014 NATIONAL SURVEY ON DRUG USE AND HEALTH (HHS PUBLICATION NO. SMA 15-4927, NSDUH SERIES H-50). ROCKVILLE, MARYLAND, USA: CENTER FOR BEHAVIORAL HEALTH STATISTICS AND QUALITY.

CITROME, L. & KETTER, T., 2013. WHEN DOES A DIFFERENCE MAKE A DIFFERENCE? INTERPRETATION OF NUMBER NEEDED TO TREAT, NUMBER NEEDED TO HARM, AND LIKELIHOOD TO BE HELPED OR HARMED. INT J CLIN PRACT, 67(5), PP. 407-411.

CITROME, L., 2010. RELATIVE VS. ABSOLUTE MEASURES OF BENEFIT AND RISK: WHAT'S THE DIFFERENCE? ACTA PSYCHIATR SCAND, 121(2), PP. 94-102.

COLWELL, C.W. JR. ET AL., 2005. ORAL DIRECT THROMBIN INHIBITOR XIMELAGATRAN COMPARED WITH WARFARIN FOR THE PREVENTION OF VENOUS THROMBOEMBOLISM AFTER TOTAL KNEE ARTHROPLASTY. J BONE JOINT SURG AM, 87(10), PP. 2169-2177.

COOK, R. & SACKETT, D., 1995. THE NUMBER NEEDED TO TREAT: A CLINICALLY USEFUL MEASURE OF TREATMENT EFFECT. BMJ, 310(6977), PP. 452-454.

CURTIN, F. & SCHULZ, P., 2011. ASSESSING THE BENEFIT: RISK RATIO OF A DRUG RANDOMIZED AND NATURALISTIC EVIDENCE. DIALOGUES CLIN NEUROSCI, 13(2), PP. 183-190.

DAHL, O.E. ET AL., 2005. POSTOPERATIVE MELAGATRAN/XIMELAGATRAN FOR THE PREVENTION OF VENOUS THROMBOEMBOLISM FOLLOWING MAJOR ELECTIVE ORTHOPAEDIC SURGERY: EFFECTS OF TIMING OF FIRST DOSE AND RISK FACTORS FOR THROMBOEMBOLISM AND BLEEDING COMPLICATIONS ON EFFICACY AND SAFETY. CLIN DRUG INVEST, 25(1), PP. 65-77.

DE ABAJO, F.J. ET AL., 2004. ACUTE AND CLINICALLY RELEVANT DRUG-INDUCED LIVER INJURY: A POPULATION BASED CASE-CONTROL STUDY. BR J CLIN PHARMACOL, 58(1), PP.71-80.

DEEKS, J., HIGGINS, J. & ALTMAN, D., 2011. CHAPTER 9: ANALYSING DATA AND UNDERTAKING META-ANALYSES. IN: T. C. COLLABORATION, ED. COCHRANE HANDBOOK FOR SYSTEMATIC REVIEWS OF INTERVENTIONS VERSION 5.1.0., S.L.: THE COCHRANE COLLABORATION.

DERSIMONIAN, R. & LAIRD, N., 1986. META-ANALYSIS IN CLINICAL TRIALS. CONTROL CLIN TRIALS. 1986, 7(3), PP. 177-188.

DERUMEAUX, G. ET AL., 2012. ECHOCARDIOGRAPHIC EVIDENCE FOR VALVULAR TOXICITY OF BENFLUOREX: A DOUBLE-BLIND RANDOMISED TRIAL IN PATIENTS WITH TYPE 2 DIABETES MELLITUS. PLOS ONE, 7(6), PP. E38273.

DESPRÉS, J.P. ET AL., 2005. EFFECTS OF RIMONABANT ON METABOLIC RISK FACTORS IN OVERWEIGHT PATIENTS WITH DYSLIPIDEMIA. N ENGL J MED, 353(20), PP. 2121-2134.

DOOREN, J. & WHALEN, J., 2007. FDA PANEL REJECTS SANOFI DRUG FOR OBESITY. THE WALL STREET JOURNAL, 14 JUNE.

DOSHI, P. & JEFFERSON, T., 2013. CLINICAL STUDY REPORTS OF RANDOMISED CONTROLLED TRIALS: AN EXPLORATORY REVIEW OF PREVIOUSLY CONFIDENTIAL INDUSTRY REPORTS. BMJ OPEN, 3(2), P. E002496.

EHRICH, E.W. ET AL., 2001. EFFECT OF ROFECOXIB THERAPY ON MEASURES OF HEALTH-RELATED QUALITY OF LIFE IN PATIENTS WITH OSTEOARTHRITIS. AM J MANAG CARE, 7(6), PP. 609-616.

EMA, 2001. SCIENTIFIC CONCLUSIONS - OVERALL SUMMARY OF THE SCIENTIFIC EVALUATION OF SIBUTRAMINE CONTAINING MEDICINAL PRODUCTS. [ONLINE] AVAILABLE AT: HTTP://EC.EUROPA.EU/HEALTH/DOCUMENTS/COMMUNITY-REGISTER/2001/200103264189/ANX_4189_EN.PDF [ACCESSED 09 FEBRUARY 2016].

EMA, 2004. EMEA STATEMENT FOLLOWING WITHDRAWAL OF VIOXX (ROFECOXIB) [PRESS RELEASE: 06-10-2004]. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/PRESS_RELEASE/2009/11/WC500015262.PDF [ACCESSED 15 SEPTEMBER 2015].

EMA, 2006A. AVANDIA EUROPEAN PUBLIC ASSESSMENT REPORT – SCIENTIFIC DISCUSSION. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/EPAR_-_SCIENTIFIC_DISCUSSION/HUMAN/000268/WC500029103.PDF [ACCESSED 03 MAY 2016].

EMA, 2006B. ACOMPLIA EUROPEAN PUBLIC ASSESSMENT REPORT – SCIENTIFIC DISCUSSION. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/EPAR_-_SCIENTIFIC_DISCUSSION/HUMAN/000666/WC500021284.PDF [ACCESSED 15 FEBRUARY 2016].

EMA, 2008A. CHMP ASSESSMENT REPORT FOR TREDAPTIVE (DOC.REF.: EMEA/348364/2008). [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/EPAR_-_PUBLIC_ASSESSMENT_REPORT/HUMAN/000889/WC500042219.PDF [ACCESSED 09 FEBRUARY 2016].

EMA, 2008B. QUESTIONS AND ANSWERS ON THE RECOMMENDATION TO SUSPEND THE MARKETING AUTHORISATION OF ACOMPLIA (RIMONABANT). [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/MEDICINE_QA/2009/11/WC500014779.PDF [ACCESSED 02 MAY 2016].

EMA, 2009A. BENEFIT-RISK METHODOLOGY PROJECT. DEVELOPMENT AND TESTING OF TOOLS AND PROCESSES FOR BALANCING MULTIPLE BENEFITS AND RISKS AS AN AID TO INFORMED REGULATORY DECISIONS ABOUT MEDICINAL PRODUCTS. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REPORT/2011/07/WC500109477.PDF [ACCESSED 15 FEB 2017].

EMA, 2010B. EUROPEAN MEDICINES AGENCY RECOMMENDS SUSPENSION OF AVANDIA, AVANDAMET AND AVAGLIM - ANTI-DIABETES MEDICATION TO BE TAKEN OFF THE MARKET [PRESS RELEASE: 23-09-2010]. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/EMA/INDEX.JSP?CURL=PAGES/NEWS_AND_EVENTS/NEWS/2010/09/NEWS_DETAIL_001119.JSP&MID=WC0B01AC058004D5C1 [ACCESSED 23 APRIL 2015].

EMA, 2010C. ASSESSMENT REPORT FOR AVANDIA (PROCEDURE NO EMA/H/C/000268/A20/0075). [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/EPAR_-_ASSESSMENT_REPORT_-_VARIATION/HUMAN/000268/WC500100757.PDF [ACCESSED 23 APRIL 2015].

EMA, 2013. ASSESSMENT REPORT FOR ALMITRINE-CONTAINING MEDICINAL PRODUCTS FOR ORAL USE. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REFERRALS_DOCUMENT/ALMITRINE/RECOMMENDATION_PROVIDED_BY_PHARMACOVIGILANCE_RISK_ASSESSMENT_COMMITTEE/WC500144134.PDF [ACCESSED 09 FEB 2016].

EMA, 2013B. EUROPEAN MEDICINES AGENCY CONFIRMS RECOMMENDATION TO SUSPEND TREDAPTIVE, PELZONT AND TREVACLYN [PRESS RELEASE: 18-01-2013]. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/EMA/INDEX.JSP?CURL=PAGES/NEWS_AND_EVENTS/NEWS/2013/01/NEWS_DETAIL_001694.JSP&MID=WC0B01AC058001D126 [ACCESSED 09 FEBRUARY 2016].

ERIKSSON, B. ET AL., 2002A. A DOSE-RANGING STUDY OF THE ORAL DIRECT THROMBIN INHIBITOR, XIMELAGATRAN, AND ITS SUBCUTANEOUS FORM, MELAGATRAN, COMPARED WITH DALTEPARIN IN

THE PROPHYLAXIS OF THROMBOEMBOLISM AFTER HIP OR KNEE REPLACEMENT: METHRO I. THROMB HAEMOST, 87(2), PP. 231-237.

ERIKSSON, B. ET AL., 2002B. XIMELAGATRAN AND MELAGATRAN COMPARED WITH DALTEPARIN FOR PREVENTION OF VENOUS THROMBOEMBOLISM AFTER TOTAL HIP OR KNEE REPLACEMENT: THE METHRO II RANDOMISED TRIAL. LANCET, 360(9344), PP. 1441-1447.

ERIKSSON, B. ET AL., 2003A. DIRECT THROMBIN INHIBITOR MELAGATRAN FOLLOWED BY ORAL XIMELAGATRAN IN COMPARISON WITH ENOXAPARIN FOR PREVENTION OF VENOUS THROMBOEMBOLISM AFTER TOTAL HIP OR KNEE REPLACEMENT. THROMB HAEMOST, 89(2), PP. 288-296.

ERIKSSON, B. ET AL., 2003B. THE DIRECT THROMBIN INHIBITOR MELAGATRAN FOLLOWED BY ORAL XIMELAGATRAN COMPARED WITH ENOXAPARIN FOR THE PREVENTION OF VENOUS THROMBOEMBOLISM AFTER TOTAL HIP OR KNEE REPLACEMENT: THE EXPRESS STUDY. J THROMB HAEMOST, 1(12), PP. 2490-2496.

ERIKSSON, B. ET AL., 2004. SIGNIFICANTLY LOWER NEED FOR BLOOD TRANSFUSIONS ASSOCIATED WITH POST-OPERATIVELY INITIATED SUBCUTANEOUS MELAGATRAN/ORAL XIMELAGATRAN COMPARED WITH ENOXAPARIN. THROMB HAEMOST, 92(2), PP. 428-430.

FDA, 2000. CDER-PHRMA-AASLD CONFERENCE 2000. CLINICAL WHITE PAPER. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DOWNLOADS/DRUGS/SCIENCERESEARCH/RESEARCHAREAS/UCM091457.PDF [ACCESSED 06 MAY 2016].

FDA, 2004. DIVISION OF CARDIOVASCULAR AND RENAL DRUG. NDA 21-686. XIMELAGATRAN (H376/95). [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/OHRMS/DOCKETS/AC/04/BRIEFING/2004-4069B1_06_FDA-BACKGROUNDER-C-R-MOR.PDF [ACCESSED 09 FEBRUARY 2016].

FDA, 2010A. FDA BRIEFING DOCUMENT: ADVISORY COMMITTEE MEETING FOR NDA 21071 AVANDIA (ROSIGLITAZONE MALEATE) TABLET (JULY 13 AND 14, 2010). [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DOWNLOADS/ADVISORYCOMMITTEES/COMMITTEESMEETINGMATERIALS/DRUGS/ENDOCRINOLOGICANDMETABOLICDRUGSADVISORYCOMMITTEE/UCM218493.PDF [ACCESSED 01 APRIL 2015].

FDA, 2011. FDA DRUG SAFETY COMMUNICATION: UPDATED RISK EVALUATION AND MITIGATION STRATEGY (REMS) TO RESTRICT ACCESS TO ROSIGLITAZONE CONTAINING MEDICINES INCLUDING AVANDIA, AVANDAMET, AND AVANDARYL [PRESS RELEASE: 18-05-2011]. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DRUGS/DRUGSAFETY/UCM255005.HTM#SAFETY_ANNOUNCEMENT [ACCESSED 23 APRIL 2015].

FDA, 2013A. STRUCTURED APPROACH TO BENEFIT-RISK ASSESSMENT IN DRUG REGULATORY DECISION-MAKING: DRAFT PATIENT DRUG USER FREE ACT (PDUFA) I IMPLEMENTATION PLAN. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DOWNLOADS/FORINDUSTRY/USERFEES/PRESCRIPTIONDRUGUSERFEE/UCM329758.PDF [ACCESSED 02 APRIL 2015].

FDA, 2013B. READJUDICATION OF THE ROSIGLITAZONE EVALUATED FOR CV OUTCOMES AND REGULATION OF GLYCEMIA IN DIABETES TRIAL (RECORD). JOINT MEETING OF THE ENDOCRINOLOGIC AND METABOLIC DRUGS ADVISORY COMMITTEE AND THE DRUG SAFETY AND RISK MANAGEMENT ADVISORY COMMITTEE. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DOWNLOADS/ADVISORYCOMMITTEES/COMMITTEESMEETINGMATERIALS/DRUGS/ENDOCRINOLOGICANDMETABOLICDRUGSADVISORYCOMMITTEE/UCM354859.PDF [ACCESSED 02 APRIL 2015].

FRANCIS, C.W. ET AL., 2003. COMPARISON OF XIMELAGATRAN WITH WARFARIN FOR THE PREVENTION OF VENOUS THROMBOEMBOLISM AFTER TOTAL KNEE REPLACEMENT. N ENGL J MED, 349(18), PP. 1703-1712.

FRANKLIN, S. & WONG, N., 2013. HYPERTENSION AND CARDIOVASCULAR DISEASE: CONTRIBUTIONS OF THE FRAMINGHAM HEART STUDY. GLOB HEART, VOLUME 81, PP. 49-57.

FURUKAWA, T., GUYATT, G. & GRIFFITH, L., 2002. CAN WE INDIVIDUALIZE THE 'NUMBER NEEDED TO TREAT'? AN EMPIRICAL STUDY OF SUMMARY EFFECT MEASURES IN META-ANALYSES. INT J EPIDEMIOL, 31(1), PP. 72-76.

GERRITS, C.M. ET AL., 2007. A COMPARISON OF PIOGLITAZONE AND ROSIGLITAZONE FOR HOSPITALIZATION FOR ACUTE MYOCARDIAL INFARCTION IN TYPE 2 DIABETES. PHARMACOEPIDEMIOL DRUG SAF, 16(10), PP. 1065-1071.

GLEIM, G. ET AL., 2009. EFFICACY AND SAFETY PROFILE OF CO-ADMINISTERED ER NIACIN/LAROPIPRANT AND SIMVASTATIN IN DYSLIPIDAEMIA. BR J CARDIOL, 16, PP. 90–97.

GOLDER, S. & LOKE, Y., 2008. IS THERE EVIDENCE FOR BIASED REPORTING OF PUBLISHED ADVERSE EFFECTS DATA IN PHARMACEUTICAL INDUSTRY-FUNDED STUDIES? BR J CLIN PHARMACOL, 66(6), PP. 767-773.

GOLDER, S., LOKE, Y. & BLAND, M., 2013. COMPARISON OF POOLED RISK ESTIMATES FOR ADVERSE EFFECTS FROM DIFFERENT OBSERVATIONAL STUDY DESIGNS: METHODOLOGICAL OVERVIEW. PLOS ONE, 8(8), P. E71813.

GÓRECKA, D. ET AL., 2003. EFFECTS OF ALMITRINE BISMESYLATE ON ARTERIAL BLOOD GASES IN PATIENTS WITH CHRONIC OBSTRUCTIVE PULMONARY DISEASE AND MODERATE HYPOXAEMIA: A MULTICENTRE, RANDOMISED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY. RESPIRATION, 70(3), PP. 275-283.

GRAHAM, D. ET AL., 2010. RISK OF ACUTE MYOCARDIAL INFARCTION, STROKE, HEART FAILURE, AND DEATH IN ELDERLY MEDICARE PATIENTS TREATED WITH ROSIGLITAZONE OR PIOGLITAZONE. JAMA, 304(4.), PP. 411-418.

GSK-CLINICALSTUDYREGISTER.COM, 2016. GSK CLINICAL STUDY REGISTER [ONLINE] AVAILABLE AT: HTTP://WWW.GSK-CLINICALSTUDYREGISTER.COM/ [ACCESSED 3 MAY 2016].

HAMMAD, T. ET AL., 2013. THE FUTURE OF POPULATION-BASED POSTMARKET DRUG RISK ASSESSMENT: A REGULATOR'S PERSPECTIVE. CLIN PHARMACOL THER, 94(3), PP. 349-358.

HAWKEY, C. ET AL., 2000. COMPARISON OF THE EFFECT OF ROFECOXIB (A CYCLOOXYGENASE 2 INHIBITOR), IBUPROFEN, AND PLACEBO ON THE GASTRODUODENAL MUCOSA OF PATIENTS WITH OSTEOARTHRITIS: A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED TRIAL. THE ROFECOXIB OSTEOARTHRITIS ENDOSCOPY MULTINATIONAL STUDY GROUP. ARTHRITIS RHEUM, 43(2), PP. 370-377.

HIGGINS, J., THOMPSON, S., DEEKS, J. & ALTMAN, D., 2003. MEASURING INCONSISTENCY IN META-ANALYSES. BMJ, 327(7414), PP. 557-560.

HOME, P. ET AL., 2009. ROSIGLITAZONE EVALUATED FOR CARDIOVASCULAR OUTCOMES IN ORAL AGENT COMBINATION THERAPY FOR TYPE 2 DIABETES (RECORD): A MULTICENTRE, RANDOMISED, OPEN-LABEL TRIAL. LANCET, 373(9681), PP. 2125-2135.

HPS2-THRIVE COLLABORATIVE GROUP, 2014. EFFECTS OF EXTENDED-RELEASE NIACIN WITH LAROPIPRANT IN HIGH-RISK PATIENTS. N ENGL J MED, 371(3), PP. 203-212.

HUGHES, S., COHEN, D. & JAGGI, R., 2014. DIFFERENCES IN REPORTING SERIOUS ADVERSE EVENTS IN INDUSTRY SPONSORED CLINICAL TRIAL REGISTRIES AND JOURNAL ARTICLES ON ANTIDEPRESSANT AND ANTIPSYCHOTIC DRUGS: A CROSS-SECTIONAL STUDY. BMJ OPEN, 4(7), P. E005535.

JAMES, W. ET AL., 2010. EFFECT OF SIBUTRAMINE ON CARDIOVASCULAR OUTCOMES IN OVERWEIGHT AND OBESE SUBJECTS. N ENGL J MED, 363(10), PP. 905-917.

JAMES, W. ET AL., 2000. EFFECT OF SIBUTRAMINE ON WEIGHT MAINTENANCE AFTER WEIGHT LOSS: A RANDOMISED TRIAL. STORM STUDY GROUP. SIBUTRAMINE TRIAL OF OBESITY REDUCTION AND MAINTENANCE. LANCET, 356(9248), PP. 2119-2125.

JEFFREY, S., 2004. FDA PANEL RECOMMENDS AGAINST APPROVAL OF ALL INDICATIONS FOR XIMELAGATRAN, S.L.: MEDSCAPE.

JÜNI, P. ET AL., 2004. RISK OF CARDIOVASCULAR EVENTS AND ROFECOXIB: CUMULATIVE META-ANALYSIS. LANCET, 364(9450), PP. 2021-2029.

JUURLINK, D.N. ET AL., 2009. ADVERSE CARDIOVASCULAR EVENTS DURING TREATMENT WITH PIOGLITAZONE AND ROSIGLITAZONE: POPULATION BASED COHORT STUDY. BMJ, 339, PP. B2942.

KAUKUA, J.K., PEKKARINEN, T.A. & RISSANEN, A.M., 2004. HEALTH-RELATED QUALITY OF LIFE IN A RANDOMISED PLACEBO-CONTROLLED TRIAL OF SIBUTRAMINE IN OBESE PATIENTS WITH TYPE II DIABETES. INT J OBES RELAT METAB DISORD, 28(4), PP. 600-605.

LAINE, L. ET AL., 1999. A RANDOMIZED TRIAL COMPARING THE EFFECT OF ROFECOXIB, A CYCLOOXYGENASE 2-SPECIFIC INHIBITOR, WITH THAT OF IBUPROFEN ON THE GASTRODUODENAL MUCOSA OF PATIENTS WITH OSTEOARTHRITIS. ROFECOXIB OSTEOARTHRITIS ENDOSCOPY STUDY GROUP. GASTROENTEROLOGY, 117(4), PP. 776-783.

LAUPACIS, A., SACKETT, D. & ROBERTS, R., 1988. AN ASSESSMENT OF CLINICALLY USEFUL MEASURES OF THE CONSEQUENCES OF TREATMENT. N ENGL J MED, 318(26), PP. 1728-1733.

LOKE, Y., PRICE, D., HERXHEIMER, A. & METHODS, A. G. C. A. E., 2007. SYSTEMATIC REVIEWS OF ADVERSE EFFECTS: FRAMEWORK FOR A STRUCTURED APPROACH. BMC MED RES METHODOL, 5(7), P. 32.

LOOK AHEAD RESEARCH GROUP, 2013. CARDIOVASCULAR EFFECTS OF INTENSIVE LIFESTYLE INTERVENTION IN TYPE 2 DIABETES. N ENGL J MED, 369(2), PP. 145-154.

MACCUBBIN, D. ET AL., 2008. LIPID-MODIFYING EFFICACY AND TOLERABILITY OF EXTENDED-RELEASE NIACIN/LAROPIPRANT IN PATIENTS WITH PRIMARY HYPERCHOLESTEROLAEMIA OR MIXED DYSLIPIDAEMIA. INT J CLIN PRACT, 62(12), PP. 1959-1970.

MACCUBBIN, D. ET AL., 2009. FLUSHING PROFILE OF EXTENDED-RELEASE NIACIN/LAROPIPRANT VERSUS GRADUALLY TITRATED NIACIN EXTENDED-RELEASE IN PATIENTS WITH DYSLIPIDEMIA WITH AND WITHOUT ISCHEMIC CARDIOVASCULAR DISEASE. AM J CARDIOL, 104(1), PP. 74-81.

MCNULTY, S.J. ET AL., 2003. A RANDOMIZED TRIAL OF SIBUTRAMINE IN THE MANAGEMENT OF OBESE TYPE 2 DIABETIC PATIENTS TREATED WITH METFORMIN. DIABETES CARE, 26(1), PP. 125-131.

MCQUAY, H. & MOORE, R., 1997. USING NUMERICAL RESULTS FROM SYSTEMATIC REVIEWS IN CLINICAL PRACTICE. ANN INTERN MED, 126(9), PP. 712-720.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2015. NUMBER NEEDED TO HARM IN THE POST-MARKETING SAFETY EVALUATION: RESULTS FOR ROSIGLITAZONE AND PIOGLITAZONE. PHARMACOEPIDEMIOL DRUG SAF, 24(12), PP. 1259-1270.

MT-ISA, S. ET AL., 2014. BALANCING BENEFIT AND RISK OF MEDICINES: A SYSTEMATIC REVIEW AND CLASSIFICATION OF AVAILABLE METHODOLOGIES. PHARMACOEPIDEMIOL DRUG SAF, 23(7), PP. 667-678.

NISSEN, S. & WOLSKI, K., 2007. EFFECT OF ROSIGLITAZONE ON THE RISK OF MYOCARDIAL INFARCTION AND DEATH FROM CARDIOVASCULAR CAUSES. N ENGL J MED, 356(24), PP. 2457-2471.

NISSEN, S. & WOLSKI, K., 2010. ROSIGLITAZONE REVISITED: AN UPDATED META-ANALYSIS OF RISK FOR MYOCARDIAL INFARCTION AND CARDIOVASCULAR MORTALITY. ARCH INTERN MED, 170(14), PP. 1191-1201.

NISSEN, S. ET AL., 2008. EFFECT OF RIMONABANT ON PROGRESSION OF ATHEROSCLEROSIS IN PATIENTS WITH ABDOMINAL OBESITY AND CORONARY ARTERY DISEASE: THE STRADIVARIUS RANDOMIZED CONTROLLED TRIAL. JAMA, 299(13), PP. 1547-1560.

NISSEN, S. ET AL., 2008. STRADIVARIUS. EFFECT OF RIMONABANT ON PROGRESSION OF ATHEROSCLEROSIS IN PATIENTS WITH ABDOMINAL OBESITY AND CORONARY ARTERY DISEASE: THE STRADIVARIUS RANDOMIZED CONTROLLED TRIAL. JAMA, 299(13), PP. 1547-1560.

PI-SUNYER, F.X. ET AL., 2006. RIO-NORTH AMERICA STUDY GROUP. EFFECT OF RIMONABANT, A CANNABINOID-1 RECEPTOR BLOCKER, ON WEIGHT AND CARDIOMETABOLIC RISK FACTORS IN OVERWEIGHT OR OBESE PATIENTS: RIO-NORTH AMERICA: A RANDOMIZED CONTROLLED TRIAL. JAMA, 295(7), PP. 761-775.

POTTHAST, R. ET AL., 2014. IMPACT OF INCLUSION OF INDUSTRY TRIAL RESULTS REGISTRIES AS AN INFORMATION SOURCE FOR SYSTEMATIC REVIEWS. PLOS ONE, 9(4), P. E92067.

SAAG, K., ET AL., 2000. ROFECOXIB, A NEW CYCLOOXYGENASE 2 INHIBITOR, SHOWS SUSTAINED EFFICACY, COMPARABLE WITH OTHER NONSTEROIDAL ANTI-INFLAMMATORY DRUGS: A 6-WEEK AND A 1-YEAR TRIAL IN PATIENTS WITH OSTEOARTHRITIS. OSTEOARTHRITIS STUDIES GROUP. ARCH FAM MED, 9(10), PP. 1124-1234.

SCHEEN, A.J. ET AL., 2006. EFFICACY AND TOLERABILITY OF RIMONABANT IN OVERWEIGHT OR OBESE PATIENTS WITH TYPE 2 DIABETES: A RANDOMISED CONTROLLED STUDY. LANCET, 368(9548), PP. 1660-1672.

SINGAL, A., HIGGINS, P. & WALJEE, A., 2014. A PRIMER ON EFFECTIVENESS AND EFFICACY TRIALS. CLIN TRANSL GASTROENTEROL, VOLUME 5, P. E45.

SINGH, S. & LOKE, Y., 2012. DRUG SAFETY ASSESSMENT IN CLINICAL TRIALS: METHODOLOGICAL CHALLENGES AND OPPORTUNITIES. TRIALS, 20(13), P. 138.

SMEETH, L., HAINES, A. & EBRAHIM, S., 1999. NUMBERS NEEDED TO TREAT DERIVED FROM META-ANALYSES–SOMETIMES INFORMATIVE, USUALLY MISLEADING. BMJ, 318(7197), PP. 1548-1551.

SMITH, I.G. ET AL., 2001. RANDOMIZED PLACEBO-CONTROLLED TRIAL OF LONG-TERM TREATMENT WITH SIBUTRAMINE IN MILD TO MODERATE OBESITY. J FAM PRACT, 50(6), PP. 505-512.

STANLEY, K., 2007. DESIGN OF RANDOMIZED CONTROLLED TRIALS. CIRCULATION, 115(9), PP. 1164-1169.

STRAUS, S., GLASZIOU, P., RICHARDSON, W. & HAYNES, R., 2011. EVIDENCE-BASED MEDICINE: HOW TO PRACTICE AND TEACH IT. 4TH ED. LONDON, UK: CHURCHILL LIVINGSTONE.

TRUITT, K.E., ET AL., 2001. A MULTICENTER, RANDOMIZED, CONTROLLED TRIAL TO EVALUATE THE SAFETY PROFILE, TOLERABILITY, AND EFFICACY OF ROFECOXIB IN ADVANCED ELDERLY PATIENTS WITH OSTEOARTHRITIS. AGING (MILANO), 13(2), PP. 112-21.

VAN GAAL, L.F. ET AL., 2005. EFFECTS OF THE CANNABINOID-1 RECEPTOR BLOCKER RIMONABANT ON WEIGHT REDUCTION AND CARDIOVASCULAR RISK FACTORS IN OVERWEIGHT PATIENTS: 1-YEAR EXPERIENCE FROM THE RIO-EUROPE STUDY. LANCET, 365(9468), PP. 1389-97.

VANDENBROUCKE, J. & PSATY, B., 2008. BENEFITS AND RISKS OF DRUG TREATMENTS: HOW TO COMBINE THE BEST EVIDENCE ON BENEFITS WITH THE BEST DATA ABOUT ADVERSE EFFECTS. JAMA, 300(20), PP. 2417-2419.

VOISIN, C., HOWARD, P. & ANSQUER, J.C., 1987. ALMITRINE BISMESYLATE: A LONG-TERM PLACEBO-CONTROLLED DOUBLE-BLIND STUDY IN COAD-VECTARION INTERNATIONAL MULTICENTRE STUDY GROUP. BULL EUR PHYSIOPATHOL RESPIR, SUPPL 11, PP. 169S-182S.

WALKER, A.M., KORO, C.E. & LANDON, J., 2008. CORONARY HEART DISEASE OUTCOMES IN PATIENTS RECEIVING ANTIDIABETIC AGENTS IN THE PHARMETRICS DATABASE 2000-2007. PHARMACOEPIDEMIOL DRUG SAF, 17(8), PP. 760-768.

WEILL, A. ET AL., 2010. BENFLUOREX AND VALVULAR HEART DISEASE: A COHORT STUDY OF A MILLION PEOPLE WITH DIABETES MELLITUS. PHARMACOEPIDEMIOL DRUG SAF, 19(12), PP. 1256-1262.

WEITZENBLUM, E. ET AL., 1992. [SEQUENTIAL ADMINISTRATION OF A REDUCED DOSE OF ALMITRINE TO PATIENTS WITH CHRONIC OBSTRUCTIVE BRONCHOPNEUMOPATHIES. A CONTROLLED MULTICENTER STUDY]. REV MAL RESPIR, 9(4), PP. 455-463.

WINKELMAYER, W.C. ET AL., 2008. COMPARISON OF CARDIOVASCULAR OUTCOMES IN ELDERLY PATIENTS WITH DIABETES WHO INITIATED ROSIGLITAZONE VS PIOGLITAZONE THERAPY. ARCH INTERN MED, 168(21), PP. 2368-2375.

WIRTH, A. & KRAUSE, J., 2001. LONG-TERM WEIGHT LOSS WITH SIBUTRAMINE: A RANDOMIZED CONTROLLED TRIAL. JAMA, 286(11), PP. 1331-1339.

ZIYADEH, N. ET AL., 2009. THE THIAZOLIDINEDIONES ROSIGLITAZONE AND PIOGLITAZONE AND THE RISK OF CORONARY HEART DISEASE: A RETROSPECTIVE COHORT STUDY USING A US HEALTH INSURANCE DATABASE. CLIN THER, 31(11), PP. 2665-2677.

ZORZELA, L., GOLDER, S., LIU, Y. & AL, E., 2014. QUALITY OF REPORTING IN SYSTEMATIC REVIEWS OF ADVERSE EVENTS: SYSTEMATIC REVIEW. BMJ, 8(348), P. F7668.

III.7. SUPPLEMENTAL DATA III

III.7.1. SUPPLEMENTAL DATA III.1 – CHARACTERISTICS OF THE INCLUDED STUDIES

Supplemental Table III. 1 – Characteristics of the included studies.

Drug /	Study				AE	Withdrawn Drug		Control		OR	I^2	Notes
Period	Ref.	Design	Duration	Interventions		Patients, N=	Events, N=	Patients, N=	Events, N=			
Almitrine					PN							
Pre-		NA				NA	NA	NA	NA	NA	-	
Post-	Arnaud et al, 1983	RCT	6 mo	Almitrine; Placebo		200	NA	NA	NA	-	-	Excluded from the meta-analysis
Post-	Voisin et al, 1987	RCT	12 mo	Almitrine; Placebo		344	50	357	8	-	-	
Post-	Bakran et al, 1990	RCT	3 mo	Almitrine; Placebo		23	NA	17	NA	-	-	Excluded from the meta-analysis
Post-	Bardsley et al, 1991	RCT	24 mo	Almitrine; Placebo		43	6	46	2	-	-	
Post-	Weitzenblum et al, 1992	RCT	12 mo	Almitrine; Placebo		65	6	37	3	-	-	
Post-	Bardsley et al, 1992	RCT	6 mo	Almitrine; Placebo		50	2	35	2	-	-	
Post-	Gorecka et al, 2003	RCT	12 mo	Almitrine; Placebo		57	7	58	4	-	-	
Post-	Pooled	RCT	-	-		559	71	533	19	2.49 (0.98 - 6.29)	58%	
Benfluorex					HVD							
Pre-		NA				NA	NA	NA	NA	NA	-	
Post-	REGULATE (2012)	RCT	12 mo	Benfluorex; Pioglitazone		310	82	305	33	2.97 (1.91 - 4.63)	-	Cases of emergent regurgitation. OR reported in the publication.
Post-	Weill A, et al. 2010	Cohort	85677 PY	Benfluorex; Non-exposition to benfluorex		43044	65	1006129	532	3.10 (2.40 - 4.00)	-	Cases of hospitalization for cardiac valvular insufficiency for any cause, mitral insufficiency, and aortic insufficiency; valvular replacement surgery for valvular insufficiency of any cause. Adjusted RR reported in the publication.

Drug / *Period*	Study: Ref.	Design	Duration	Interventions	AE	Withdrawn Drug: Patients, N=	Withdrawn Drug: Events, N=	Control: Patients, N=	Control: Events, N=	OR	I^2	Notes
Nicotinic acid / laropiprant					MP							
Pre-	P020-02	RCT	24 wk	ERN/LRPT; ERN; Placebo		798	0	811	0	-	-	
Pre-	P022-02	RCT	12 wk	ERN/LRPT + Simvastatin; ERN/LRPT; Simvastatin		804	0	593	0	-	-	
Pre-	P054	RCT	16 wk	ERN/LRPT; ERN		725	0	727	1	-	-	
Pre-	Pooled	RCT		-		2327	0	2131	1	0.33 (0.014 - 8.21)	0%	Comparisons: ERN/LRPT ± Simvastatin Vs. ERN ± Placebo OR Simvastatin
Post-	HPS2-THRIVE	RCT	3,6 yrs	ERN/LRPT; Placebo		12838	75	12835	17	4.43 (2.62 - 7.51)	0%	Estimated OR.
Nicotinic acid / laropiprant					NOD							
Pre-	P020-02	RCT	24 wk	ERN/LRPT; ERN; Placebo		798	5	811	2	-	-	
Pre-	P022-02	RCT	12 wk	ERN/LRPT + Simvastatin; ERN/LRPT; Simvastatin		804	2	593	1	-	-	
Pre-	P054	RCT	16 wk	ERN/LRPT; ERN		725	2	727	1	-	-	
Pre-	Pooled	RCT		-		2327	9	2131	4	2.11 (0.65 - 6.87)	0%	Comparisons: ERN/LRPT ± Simvastatin Vs. ERN ± Placebo OR Simvastatin
Post-	HPS2-THRIVE	RCT	3,6 yrs	ERN/LRPT; Placebo		12838	494	12835	376	1.32 (1.16 – 1.51)	-	Rate ratio from the publication.
Nicotinic acid / laropiprant					SB							
Pre-	P020-02	RCT	24 wk	ERN/LRPT; ERN; Placebo		798	NA	811	NA	-	-	
Pre-	P022-02	RCT	12 wk	ERN/LRPT + Simvastatin; ERN/LRPT; Simvastatin		804	NA	593	NA	-	-	
Pre-	P054	RCT	16 wk	ERN/LRPT; ERN		725	NA	727	NA	-	-	
Pre-	Pooled	RCT		-		2327	NA	2131	NA	-		
Post-	HPS2-THRIVE	RCT	3,6 yrs	ERN/LRPT; Placebo		12838	326	12835	238	1.38 (1.18 – 1.62)	-	Rate ratio from the publication.
Nicotinic acid / laropiprant					MP							
Pre-	P020-02	RCT	24 wk	ERN/LRPT; ERN; Placebo		798	0	811	0	-	-	

Drug /	Study				AE	Withdrawn Drug		Control		OR	I^2	Notes
Period	Ref.	Design	Duration	Interventions		Patients, N=	Events, N=	Patients, N=	Events, N=			
Nicotinic acid / laropiprant					MP							
Pre-	P022-02	RCT	12 wk	ERN/LRPT + Simvastatin; ERN/LRPT; Simvastatin		804	0	593	0	-	-	
Pre-	P054	RCT	16 wk	ERN/LRPT; ERN		725	0	727	1	-	-	
Pre-	Pooled	RCT		-		2327	0	2131	1	0.33 (0.014 - 8.21)	0%	Comparisons: ERN/LRPT ± Simvastatin Vs. ERN ± Placebo OR Simvastatin
Post-	HPS2-THRIVE	RCT	3,6 yrs	ERN/LRPT; Placebo		12838	75	12835	17	4.43 (2.62 - 7.51)	0%	Estimated OR.
Nicotinic acid / laropiprant					NOD							
Pre-	P020-02	RCT	24 wk	ERN/LRPT; ERN; Placebo		798	5	811	2	-	-	
Pre-	P022-02	RCT	12 wk	ERN/LRPT + Simvastatin; ERN/LRPT; Simvastatin		804	2	593	1	-	-	
Pre-	P054	RCT	16 wk	ERN/LRPT; ERN		725	2	727	1	-	-	
Pre-	Pooled	RCT		-		2327	9	2131	4	2.11 (0.65 - 6.87)	0%	Comparisons: ERN/LRPT ± Simvastatin Vs. ERN ± Placebo OR Simvastatin
Post-	HPS2-THRIVE	RCT	3,6 yrs	ERN/LRPT; Placebo		12838	494	12835	376	1.32 (1.16 – 1.51)	-	Rate ratio from the publication.
Nicotinic acid / laropiprant					SB							
Pre-	P020-02	RCT	24 wk	ERN/LRPT; ERN; Placebo		798	NA	811	NA	-	-	
Pre-	P022-02	RCT	12 wk	ERN/LRPT + Simvastatin; ERN/LRPT; Simvastatin		804	NA	593	NA	-	-	
Pre-	P054	RCT	16 wk	ERN/LRPT; ERN		725	NA	727	NA	-	-	
Pre-	Pooled	RCT		-		2327	NA	2131	NA	-		
Post-	HPS2-THRIVE	RCT	3,6 yrs	ERN/LRPT; Placebo		12838	326	12835	238	1.38 (1.18 – 1.62)	-	Rate ratio from the publication.
Nicotinic acid / laropiprant					SI							
Pre-	P020-02	RCT	24 wk	ERN/LRPT; ERN; Placebo		798	NA	811	NA	-	-	
Pre-	P022-02	RCT	12 wk	ERN/LRPT + Simvastatin; ERN/LRPT; Simvastatin		804	NA	593	NA	-	-	

Drug / *Period*	Study				AE	Withdrawn Drug		Control		OR	I^2	Notes
	Ref.	Design	Duration	Interventions		Patients, N=	Events, N=	Patients, N=	Events, N=			
Nicotinic acid / laropiprant					SI							
Pre-	P054	RCT	16 wk	ERN/LRPT; ERN		725	NA	727	NA	-	-	
Pre-	Pooled	RCT		-		2327	NA	2131	NA	-	-	
Post-	HPS2-THRIVE	RCT	3,6 yrs	ERN/LRPT; Placebo		12838	1031	12835	853	1.22 (1.12 – 1.34)	-	Rate ratio from the publication.
Rimonabant					PD							
Pre-	RIO EUROPE	RCT	104 wk	Rimonabant; Placebo		599	42	305	16	-	-	
Pre-	RIO LIPIDS	RCT	52 wk	Rimonabant; Placebo		346	26	342	8	-	-	
Pre-	RIO NORTH AMERICA	RCT	52 wk	Rimonabant; Placebo		1219	76	607	14	-	-	
Pre-	RIO DIABETES	RCT	52 wk	Rimonabant; Placebo		339	13	348	3	-	-	
Pre-	Pooled	RCT	-	-		2503	157	1602	41	2.48 (1.49 - 4.12)	46%	PD leading to treatment discontinuation. Estimated OR.
Post-	STRADIVA-RIUS	RCT	18 mo	Rimonabant; Placebo		422	40	416	13	3.25 (1.71 - 6.16)	-	PD leading to treatment discontinuation. Estimated OR.
Rofecoxib					MI							
Pre-	010	RCT	6 wk	Rofecoxib; Placebo	MI	147	NR	72	NR	-	-	Excluded from the meta-analysis
Pre-	029	RCT	6 wk	Rofecoxib; Placebo		527	3	145	0	-	-	
Pre-	033	RCT	6 wk	Rofecoxib; Ibuprofen; Placebo		446	NR	290	NR	-	-	Excluded from the meta-analysis
Pre-	034	RCT	52 wk	Rofecoxib; Diclofenac		463	2	230	1	-	-	
Pre-	040	RCT	6 wk	Rofecoxib; Ibuprofen; Placebo		486	NR	323	NR	-	-	Excluded from the meta-analysis
Pre-	058	RCT	6 wk	Rofecoxib; Nabutenone; Placebo		174	0	167	1	-	-	
Pre-	044	RCT	24 wk	Rofecoxib; Ibuprofen; Placebo		381	0	361	1	-	-	
Pre-	045	RCT	24 wk	Rofecoxib; Ibuprofen; Placebo		388	0	387	1	-	-	
Pre-	035	RCT	52 wk	Rofecoxib; Diclofenac		516	2	268	1	-	-	

Drug /	Study				AE	Withdrawn Drug		Control		OR	I^2	Notes
Period	Ref.	Design	Duration	Interventions		Patients, N=	Events, N=	Patients, N=	Events, N=			
Rofecoxib					MI							
Pre-	Pooled	RCT	-	-		2449	7	1558	5	0.72 (0.23 - 2.27)	0%	Estimated OR.
Post-	APPROVe	RCT	3 yr	Rofecoxib; Placebo		1287	21	1299	9	2.38 (1.09 - 5.21)	0%	Estimated OR.
Rosiglitazone					MI							
	49653/006	RCT	12 wk	Rosiglitazone; Placebo		305	1	75	0	-	-	
	49653/011	RCT	24 wk	Rosiglitazone; Placebo		357	2	176	0	-	-	
	49653/015	RCT	24 wk	Rosiglitazone and sulfonylurea; Sulfonylurea		395	2	198	1	-	-	
	49653/020	RCT	52 wk	Rosiglitazone; Glyburide		391	2	207	1	-	-	
	49653/024	RCT	26 wk	Rosiglitazone; Placebo		774	1	185	1	-	-	
	49653/079	RCT	26 wk	Rosiglitazone ± glyburide; Glyburide		203	1	106	1	-	-	
	49653/090	RCT	8 wk	Rosiglitazone; Placebo		228	1	75	0	-	-	
	49653/093	RCT	26 wk	Rosiglitazone ± metformin; Metformin		213	0	109	1	-	-	
	49653/094	RCT	26 wk	Rosiglitazone and metformin; Metformin		232	1	116	0	-	-	
	49653/096	RCT	26 wk	Rosiglitazone and glyburide; Glyburide		232	0	115	0	-	-	
	49653/098	RCT	8 wk	Rosiglitazone; Placebo		284	1	96	0	-	-	
Pre-	Pooled	RCT	-	-		3614	12	1458	5	0.77 (0.31 - 1.91)	0.8%	Estimated OR.
Post-	FDA 2010 briefing document	FDA Meta-Analysis of RCTs	52 trials	Rosiglitazone vs. non-TZD comparator		10039	45	6956	20	1.80 (1.03 - 3.25)	-	
Post-	Nissen and Wolski 2010	Nissen and Wolski's Meta-Analysis of RCTs	56 trials	Rosiglitazone vs. non-TZD comparator		17258	159	14449	136	1.28 (1.02 - 1.63)	-	

Drug / *Period*	Study				AE	Withdrawn Drug		Control		OR	I^2	Notes
	Ref.	Design	Duration	Interventions		Patients, N=	Events, N=	Patients, N=	Events, N=			
Rosiglitazone					MI							
Post-	RECORD	RCT	5.5 yr	Rosiglitazone + metformin or sulfonylurea; combination of metformin and sulfonylurea		2220	64	2227	56	1.14 (0.80 - 1.63)	-	
Post-	Brownstein et al. 2010	Cohort	353 PY	Rosiglitazone; pioglitazone		1879	133	806	36	-	-	
Post-	Ziyadeh et al. 2009	Cohort	8,4 mo	Rosiglitazone; pioglitazone		47501	194	47501	141	-	-	
Post-	Juurlink et al. 2009	Cohort	292 days	Rosiglitazone; pioglitazone		22785	425	16951	273	-	-	
Post-	Winkelmayer et al. 2008	Cohort	215 days	Rosiglitazone; pioglitazone		14101	374	14260	363	-	-	
Post-	Walker et al. 2008	Cohort	12-18 mo	Rosiglitazone; pioglitazone		57000	180	51000	139	-	-	
Post-	Gerrits et al. 2007	Cohort	1,3 yrs	Rosiglitazone; pioglitazone		15104	214	14807	161	-	-	
Post-	Pooled	Systematic Review of Cohort Studies	-	Rosiglitazone; pioglitazone		187887	1788	168957	1325	1.21 (1.09 – 1.35)	41%	Estimated OR.
Post-	Graham et al. 2010	Cohort	5400 PY	Rosiglitazone; pioglitazone		677593	2593	159978	5386	1.18 (1.12 – 1.23)	-	PY on rosiglitazone. Adjusted HR from the publication for composite outcome of MI, stroke, heart failure, or all-cause mortality.
Sibutramine					CV event							
Pre-	SB1047	RCT	52 wk	Sibutramine; Placebo		216	0	163	0	-	-	
Pre-	SB1048	RCT	104 wk	Sibutramine; Placebo		350	0	114	0	-	-	
Pre-	SB1049	RCT	52 wk	Sibutramine; Placebo		82	0	78	0	-	-	
Pre-	KD9618	RCT	48 wk	Sibutramine; Placebo		405	0	201	0	-	-	
Pre-	SB5078	RCT	52 wk	Sibutramine; Placebo		114	0	122	0	-	-	
Pre-	SB6085	RCT	52 wk	Sibutramine; Placebo		130	0	64	0	-	-	
Pre-	Pooled	RCT	-	-		1297	0	742	0	-	-	

Drug /	Study				AE	Withdrawn Drug		Control		OR	I^2	Notes
Period	Ref.	Design	Duration	Interventions		Patients, N=	Events, N=	Patients, N=	Events, N=			
Sibutramine					CV event							
Post-	SCOUT	RCT	3.4 yr	Sibutramine; Placebo		4906	561	4898	490	1.16 (1.03 - 1.31)	-	Adjusted HR from the publication for primary outcome event (nonfatal MI, nonfatal stroke, resuscitation after cardiac arrest, or CV death)
Ximelagatran					DILI							
	METHRO I	RCT	6-9 days	Ximelagatran; Enoxaparin	DILI	104	0	33	0	-	-	
	METHRO II	RCT	7-10 days	Ximelagatran; Dalteparin		1495	0	381	0	-	-	
	METHRO III	RCT	8-11 days	Ximelagatran; Enoxaparin		1439	0	1435	0	-	-	
	EXPRESS	RCT	8-11 days	Ximelagatran; Enoxaparin		1403	0	1418	0	-	-	
	EXULT A	RCT	7-12 days	Ximelagatran; Warfarin		1537	0	764	0	-	-	
	EXULT B	RCT	7-12 days	Ximelagatran; Warfarin		1152	0	1151	0	-	-	
Pre-	Pooled	RCT	-	-		7130	0	5182	0	-	-	
Post-	EXTEND	RCT	32-38 days	Ximelagatran; Enoxaparin		557	1	601	0	8.00 (0.16 - 404.14)	-	

AE, adverse event; CV, cardiovascular; DILI, drug-induced liver injury; ERN, nicotinic acid; Ev, event; FDA, Food and Drug Administration; HVD, heart valve disease; HR, Hazard Ratio; LRPT, laropiprant; MI, Myocardial infarction; MP, myopathy; NA, not available; NNTB, number needed to treat to benefit; NNTH, number needed to treat to be harmed; NOD, New-onset diabetes; NR, not reported; OR, odds ratio; PD, Psychiatric disorders; Post-, post-marketing; PN, Peripheral neuropathy; Pre-, pre-marketing; Pts, patients; PY, person-years; RCT, randomized controlled trial; Ref, Reference; RR, Relative Risk; SB, Serious bleeding; SI, Serious infection; TZD, thiazolidinedione; wk, weeks; yr, years.

Note: *Bibliographic references for the studies cited in Supplemental Table III. 1 are presented in III.6. REFERENCES.*

CHAPTER IV – BENEFIT-RISK OF THERAPIES FOR RELAPSING-REMITTING MULTIPLE SCLEROSIS: TESTING THE NUMBER NEEDED TO TREAT TO BENEFIT (NNTB), NUMBER NEEDED TO TREAT TO HARM (NNTH) AND THE LIKELIHOOD TO BE HELPED OR HARMED (LHH): A SYSTEMATIC REVIEW AND META-ANALYSIS

IV. BENEFIT-RISK OF THERAPIES FOR RELAPSING-REMITTING MULTIPLE SCLEROSIS: TESTING THE NUMBER NEEDED TO TREAT TO BENEFIT (NNTB), NUMBER NEEDED TO TREAT TO HARM (NNTH) AND THE LIKELIHOOD TO BE HELPED OR HARMED (LHH): A SYSTEMATIC REVIEW AND META-ANALYSIS

IV.1. ABSTRACT

This study aimed to test the number needed to treat to benefit (NNTB) and to harm (NNTH), and the likelihood to be helped or harmed (LHH) when assessing benefits, risks, and benefit–risk ratios of disease-modifying therapies (DMTs) approved for relapsing–remitting multiple sclerosis (RRMS). In May 2016, we conducted a systematic review using the PubMed and Cochrane Central Register of Controlled Trials databases to identify phase III, randomized controlled trials with a duration of ≥2 years that assessed first-line (dimethyl fumarate [DMF], glatiramer acetate [GA], interferons [IFN], and teriflunomide) or second-line (alemtuzumab, fingolimod, and natalizumab) DMTs in patients with RRMS. Meta-analyses were performed to estimate relative risks (RRs) on annualized relapse rate (ARR), proportion of relapse-free patients (PPR-F), disability progression (PP-F-CDPS3M), and safety outcomes. NNTB and NNTH values were calculated applying RRs to control event rates. LHH was calculated as NNTH/NNTB ratio. The lowest NNTBs on ARR, PPR-F, and PP-FCDPS3M were found with IFN-ß-1a-SC (NNTB 3, 95% CI 2–4; NNTB 7, 95% CI 4–18; NNTB 4, 95% CI 3–7, respectively) and natalizumab (NNTB 2, 95% CI 2–3; NNTB 4, 95% CI 3–6; NNTB 9, 95% CI 6–19, respectively). The lowest NNTH on adverse events leading to treatment discontinuation was found with IFN-ß-1b (NNTH 14, 95% 2–426) versus placebo; a protective effect was noted with alemtuzumab versus IFN-ß-1a-SC (NNTB 22, 95% 17–41). LHHs >1 were more frequent with IFN-ß-1a-SC and natalizumab. These metrics may be valuable for benefit-risk assessments, as they reflect baseline risks and are easily interpreted. Before making treatment decisions, clinicians must acknowledge that a higher RR reduction with drug A as compared with drug B (versus a common comparator in trial A and trial B, respectively) does not necessarily mean that the number of patients needed to be treated for one patient to encounter one additional outcome of interest over a defined period of time is lower with drug A than with drug B. Overall, IFN-ß-1a-SC and natalizumab seem to have the most favourable benefit–risk ratios among first- and second-line DMTs, respectively.

IV.2. INTRODUCTION

Until a decade ago, the therapeutic armamentarium for relapsing–remitting multiple sclerosis (RRMS) was limited, comprising only glatiramer acetate (GA) and interferons (IFNs) (English & Aloi 2015). During recent years, however, several disease-modifying therapies (DMTs) have been approved, with different efficacy and safety profiles (Ingwersen, Aktas & Hartung 2016). Given the current therapeutic arsenal, clinicians are facing challenging decisions when prescribing DMTs for RRMS patients.

Although the benefit–risk ratio of drugs can be assessed in a qualitative manner, there are quantitative metrics that can bring objectivity and reproducibility to the process (EMA 2007), (FDA 2013a), (Yuan, Levitan & Berlin 2011), (Mendes, Alves & Batel-Marques 2015), (Mt-Isa et al. 2014). The 'number needed to treat' (NNT), for example, is easily interpreted, by telling clinicians how many patients are needed to treat with one therapy versus another in order for one patient to encounter one additional outcome of interest over a defined period of time (Laupacis, Sackett & Roberts 1988), (Cook & Sackett 1995). The notation 'number needed to treat to benefit' (NNTB) is used for beneficial outcomes; while 'number needed to treat to be harmed' (NNTH) is used for harmful outcomes (Altman 1998). Further, the ratio NNTH/NNTB, named the likelihood to be helped or harmed (LHH), can be calculated to illustrate trade-offs between benefits and harms and to inform clinicians about how many patients might benefit from treatment for each one who experiences a harmful event. In case of LHH >1, the expected benefits outweigh possible harms (Citrome & Ketter 2013).

The aim of this study was to use the NNTB, NNTH, and LHH to assess the benefits, risks, and benefit–risk ratios of DMTs that have been approved for the treatment of patients with RRMS, and to provide information on the clinical use of these DMTs.

IV.3. METHODS

The present study conforms to standard guidelines and was written according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al. 2009).

IV.3.1. LITERATURE SEARCH

PubMed and Cochrane Central Register of Controlled Trials (CENTRAL) databases were searched (until May 10, 2016) to identify studies evaluating the efficacy and safety of DMTs in patients with RRMS. Bibliographic reference lists of all relevant studies, meta-analyses, and systematic reviews were hand searched to identify additional eligible articles. The electronic databases search strategy is available in Supplemental Data IV.7.1. – Supplemental Table IV. 1.

IV.3.2. STUDY SELECTION

Titles and abstracts of all retrieved citations were screened by two independent reviewers to identify potentially relevant publications. Full texts were retrieved for relevant citations. Discrepancies were resolved by majority decision (two of three) involving a third investigator. Phase III randomized controlled trials (RCTs) with ≥2 years of duration, including RRMS patients, and assessing monotherapy with approved DMTs, were considered. The inclusion and exclusion criteria are available in Supplemental Data IV.7.2. – Supplemental Table IV. 2.

IV.3.3. DATA EXTRACTION

Data elements extracted included study design, study duration, patient population characteristics [mean age, duration of disease, Expanded Disability Status Scale (EDSS) score at baseline, and prior use of DMT], interventions, comparators, primary outcomes measures, and sources of funding.

Efficacy outcomes of interest were annualized relapse rate (ARR), proportion of patients remaining relapse-free (PPR-F), and proportion of patients remaining free of confirmed disability progression sustained for 3 months (PP-FCDPS3M), as measured at 2 years from study initiation.

The main safety outcomes of interest were extracted based on 2-year data and included the following: (1) any serious adverse event (SAE); (2) any adverse event (AE) leading to discontinuation of study drug (AELD). These safety outcomes were analysed with regards to all DMTs included in this study.

In addition, a pool of AEs of interest was established for each DMT based on the information provided in 'special warnings and precautions for use', and 'undesirable effects' from European Summaries of Products Characteristics (SPCs). Supplemental Table IV. 3 (see Supplemental Data IV.7.3.) provides an overview of the approved therapeutic indications, contraindications, special warnings and precautions for use, and most common AEs with each DMT. The pool of AEs of interest established for each DMT is presented in Supplemental Table IV. 4 (see Supplemental Data IV.7.4.).

IV.3.4. ASSESSMENT OF RISK OF BIAS IN SELECTED STUDIES

A qualitative analysis assessed each study for quality by considering features that could introduce bias, according to the Cochrane Collaboration criteria (Higgins et al. 2003). These included random sequence generation, allocation concealment, blinding of participants, blinding of outcome assessor, incomplete outcome data, selective outcome reporting, and other potential risks of bias. The risk of bias on each criterion was judged and classified as 'low', 'high', or 'unclear' (Tramacere et al. 2015). Allocation concealment, blinding of outcome assessor, and incomplete outcome data were considered to summarize the overall quality of evidence (Tramacere et al. 2015).

IV.3.5. DATA ANALYSIS AND STATISTICAL METHODS

DMTs were categorized according to therapeutic indications approved in SPCs: (1) first-line (dimethyl fumarate [DMF], GA, ß-IFNs, and teriflunomide); and (2) second-line or highly active RRMS (alemtuzumab, fingolimod, and natalizumab).

Meta-analyses were performed to pool evidence from RCTs and to estimate relative risks (RR) with 95% confidence intervals (CI). Number of patients with outcome of interest and number of randomized patients were used for estimating RR for all dichotomous outcomes. In the case of ARR, number of observed relapses and total person-years of exposure were used to compute estimates. A random-effects model was used, as it is more conservative than a fixed effects model in the presence of between-studies heterogeneity (DerSimonian & Laird 1986). Random-effects model RR was also recommended as the best summary measure for clinicians who need to calculate a patient's expected event rate (PEER)-adjusted NNT values (Furukawa, Guyatt & Griffith 2002). Between-studies heterogeneity was assessed using the I^2 measure of inconsistency (Higgins et al. 2003).

NNT values (NNTB for benefits and NNTH for harms) were then calculated by applying pooled RR (and 95% CI) from meta-analyses (or individual studies when applicable) to control event rates (CERs) [18]. Two-year CERs were obtained from control groups of RCTs assessing each DMT. Further, an assumed CER of 0.1% was used to allow calculating NNT for cases if no events had been reported in control groups. The formula used to calculate NNT values from RR results was the following: NNT = 1 /[(RR-1) x (CER)] (Straus et al. 2011). NNT values were rounded up to next whole numbers. When the treatment effect is not statistically significant at the p threshold of <0.05, the 95% CI for the RR will include unity, and the 95% CI for the NNT will include infinity (∞). In such cases, one of the confidence limits indicates benefit (NNTB) and the other will indicate harm (NNTH), with the scale for NNT going from NNTB = 1 to NNTH = 1 via infinity (Altman 1998).

LHH, the ratio of NNTH to NNTB, was calculated to determine the benefit–risk ratio of each DMT. LHH values were not calculated for situations in which point-estimate values of NNTB or NNTH values were negative.

Data analyses were performed using Stata version 12 (StataCorp, College Station, TX, USA).

IV.4. RESULTS

IV.4.1. INCLUDED STUDIES

Figure IV. 1 presents the search strategy flowchart. The search returned 1782 citations. After excluding duplicates and studies with inadequate design, 13 phase III RCTs, from 15 publications, were selected (Polman et al. 2006), (O'Connor et al. 2009), (Cohen et al. 2012), (Coles et al. 2012), (Fox et al. 2012), (Johnson et al. 1995), (Gold et al. 2012), (Kappos et al. 2010), (Calabresi et al. 2014), (Paty & Li 2001), (Paty & Li 1993), ([No authors listed] 1993), (Jacobs et al. 1996), ([No authors listed] 1998), (O'Connor et al. 2011). The main characteristics of the studies are presented in Table IV. 1. One RCT compared GA with IFN-ß-1b (O'Connor et al. 2009) and two RCTs compared alemtuzumab with IFN-ß-1a-SC (Cohen et al. 2012), (Coles et al. 2012). The remaining RCTs were controlled with placebo.

IV.4.2. RISK OF BIAS IN SELECTED STUDIES

The risk of bias of the studies is summarized in Table IV. 2 and Supplemental Figure IV. 1 (Supplemental Data IV.7.5.). The overall risk of bias was judged to be 'low' in two out of 13 (15 %) RCTs (Polman et al. 2006), ([No authors listed] 1998), 'moderate' in four (31 %) RCTs (O'Connor et al. 2009), (Johnson et al. 1995), (Paty & Li 2001), (Paty & Li 2001), ([No authors listed] 1993), (Jacobs et al. 1996), and 'high' in seven (54 %) RCTs (Cohen et al. 2012), (Coles et al. 2012), (Fox et al. 2012), (Gold et al. 2012), (Kappos et al. 2010), (Calabresi et al. 2014), (O'Connor et al. 2011).

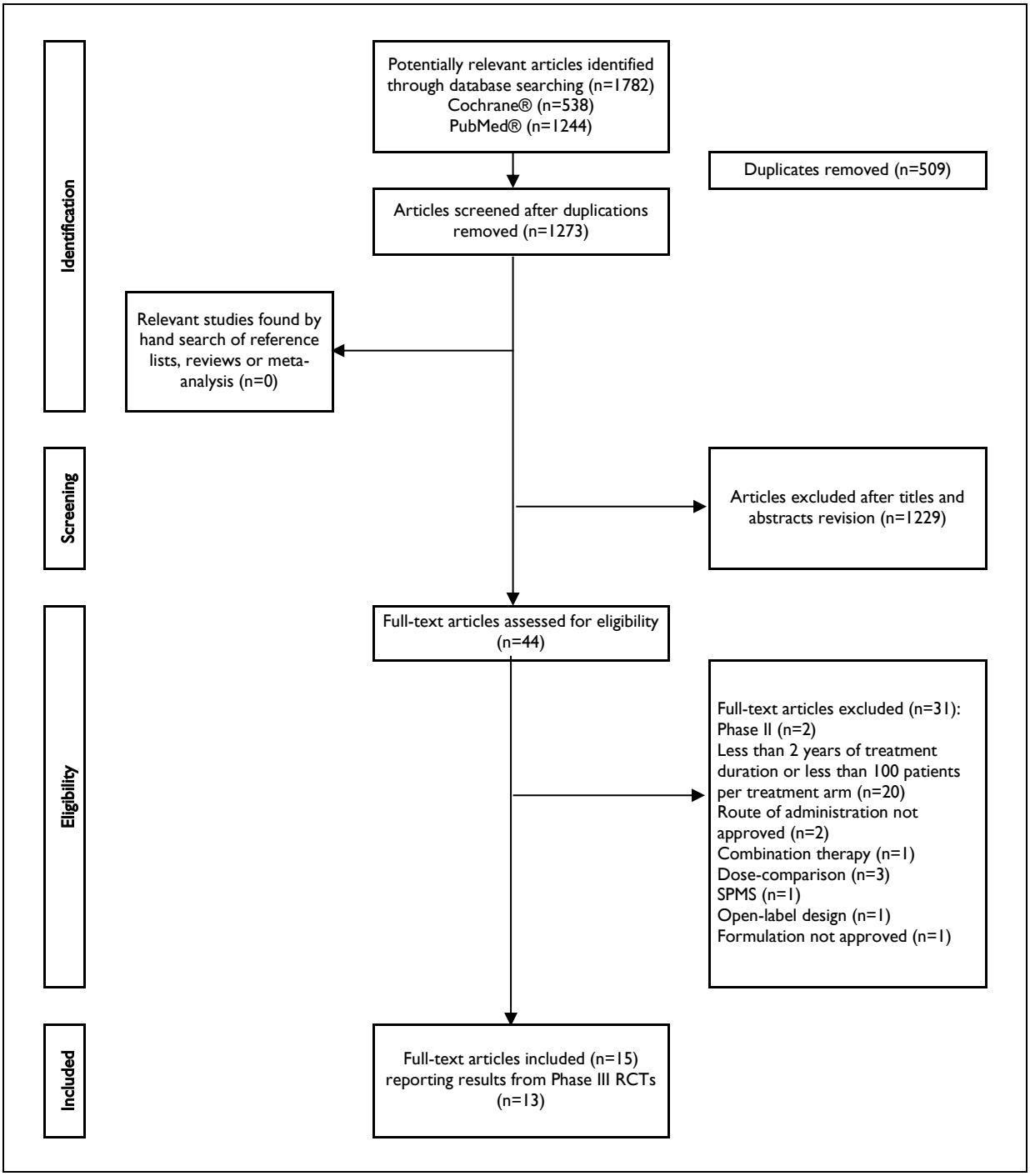


Figure IV. 1 – Flow of studies through the systematic review process.

RCT, randomized controlled trial; SPMS, secondary progressive multiple sclerosis.

Table IV. 1 – Characteristics of included studies and patients.

Study	Study type	Number of patients†	Patient characteristics	Intervention	Comparison	Length of follow-up	Primary outcome measure	Source of funding
AFFIRM 2006 **(Polman et al. 2006)**	RCT	942	Age: 18 to 50 years; diagnosis: RRMS; median disease duration: 5 years (range, 0 to 34 years); mean EDSS: 2.3; prior use of DMT: not reported	Natalizumab 300 mg IV q4w (n = 627)	Placebo (n= 315)	2 years	CDPS3M, PPR	Biogen Idec, Inc. and Elan Pharmaceuticals
BEYOND 2009 **(O'Connor et al. 2009)**	RCT	2244	Age: 18 to 55 years; diagnosis: RRMS; mean disease duration 5.3 years; mean EDSS 2.3; prior use of DMT: treatment-naïve patients	GA 20 mg SC QD (n=448)	IFNβ -1b 250 mcg SC EOD (n=897); *IFNβ -1b 500 mcg SC EOD (n=899)	2 years	PPR	Bayer HealthCare Pharmaceuticals
CARE-MS I 2012 **(Cohen et al. 2012)**	RCT	563	Age: 18 to 50 years; diagnosis: RRMS; mean disease duration: 2 years; mean EDSS 2.0; prior use of DMT: treatment-naïve patients	Alemtuzumab 12 mg IV daily for 5 days at 0 months, daily for 3 days at 12 months (n = 376)	IFNβ -1a 44 mcg SC 3 times weekly (n = 187)	2 years	CDPS6M, PPR	Genzyme (a Sanofi company)
CARE-MS II 2012 **(Coles et al. 2012)**	RCT	840	Age: 18 to 55 years; diagnosis: RRMS; mean disease duration: 5 years; mean EDSS: 2.7; prior use of DMT: all patients were previously treated (at least one relapse while on interferon β or glatiramer after at least 6 months of treatment)	Alemtuzumab 12 mg IV daily for 5 days at 0 months, daily for 3 days at 12 months (n = 436) *Alemtuzumab 24 mg IV daily for 5 days at 0 months, daily for 3 days at 12 months (n = 173)	IFNβ -1a 44 mcg SC 3 times weekly (n = 231)	2 years	CDPS6M, PPR	Genzyme (a Sanofi company)
CONFIRM 2012 **(Fox et al. 2012)**	RCT	1417	Age: 18 to 55 years; diagnosis: RRMS; mean disease duration: 5 years; mean EDSS: 2.6; prior use of DMT: 40%-41% across study groups	DMF 240 mg oral BID (n = 359); *DMF 240 mg oral TID (n = 345)	GA 20 mg SC QD (reference arm, not head-to head comparator) (n = 350); Placebo (n = 363)	2 years	ARR	Biogen Idec
Copolymer 1 MS Group 1995 **(Johnson et al. 1995)**	RCT	251	Age: 18 to 45 years; diagnosis: RRMS; mean disease duration: 7 years; mean EDSS: 2.6; prior use of DMT: not reported	GA 20 mg SC QD (n = 125)	Placebo (n = 126)	2 years	ARR	Teva Pharmaceutical
DEFINE 2012 **(Gold et al. 2012)**	RCT	1234	Age: 18 to 55 years; diagnosis: RRMS; mean disease duration: 6 years; mean EDSS: 2.4; prior use of DMT: 40%-42% across study groups	DMF 240 mg oral BID (n = 410); *DMF 240 mg oral TID (n = 416)	Placebo (n = 408)	2 years	PPR	Biogen Idec
FREEDOMS 2010 **(Kappos et al. 2010)**	RCT	1272	Age: 18 to 55 years; diagnosis: RRMS; mean disease duration: 8 years; mean EDSS: 2.4; prior use of DMT: 40%-43% across study groups	Fingolimod 0.5 mg oral QD (n = 425); *Fingolimod 1.25 mg oral QD (n = 429)	Placebo (n = 418)	2 years	ARR	Novartis Pharma
FREEDOMS II 2014 **(Calabresi et al. 2014)**	RCT	1083	Age: 18 to 55 years; diagnosis: RRMS; mean disease duration: 11 years; mean EDSS: 2.4; prior	Fingolimod 0.5 mg oral QD (n = 358); *Fingolimod 1.25 mg oral	Placebo (n = 355)	2 years	ARR	Novartis Pharma

Study	Study type	Number of patients†	Patient characteristics	Intervention	Comparison	Length of follow-up	Primary outcome measure	Source of funding
			use of DMT: 73%-78% across study groups	QD (n = 370)				
IFNB MS Group 1993 **(Paty & Li 1993)**, **(Paty & Li 2001)**, **([No authors listed] 1993)**	RCT	372	Age: 18 to 50 years; diagnosis: RRMS; mean disease duration: 4 years; mean EDSS: 2.9; prior use of DMT: not reported	IFNβ -1b 250 mcg SC EOD (n = 124); *IFNβ -1b 50 mcg SC EOD (n = 125)	Placebo (n = 123)	2 years	ARR, PPR-F	Triton Biosciences, Inc., Alameda, CA and Berlex Laboratories Inc
MSCRG 1996 **(Jacobs et al. 1996)**	RCT	301	Age: 18 to 55 years; diagnosis: RRMS; mean disease duration: 7 years; mean EDSS: 2.4; prior use of DMT: treatment-naïve patients	IFNβ -1a 30 mcg IM once weekly (n = 158)	Placebo (n = 143)	2 years	CDPS6M	Biogen, Inc, Cambridge, MA
PRISMS 1998 **([No authors listed] 1998)**	RCT	560	Age: 18 to 50 years; diagnosis: RRMS; mean disease duration: 7 years; mean EDSS: 2.5; prior use of DMT: 3% of patients had received previous immunosuppressive therapy	*IFNβ -1a 22 mcg SC TIW (n = 184); IFN β -1a 44 mcg SC TIW (n = 189)	Placebo (n = 187)	2 years	PPR	Ares-Serono International SA, Geneva, Switzerland
TEMSO 2011 **(O'Connor et al. 2011)**	RCT	1088	Age: 18 to 55 years; diagnosis: RRMS; mean disease duration: 9 years; mean EDSS: 2.7; prior use of DMT: 25%-28% across study groups	*Teriflunomide 7 mg oral QD (n = 366); Teriflunomide 14 mg oral QD (n = 359)	Placebo (n = 363)	2 years	ARR	Sanofi-Aventis

ARR: Annualized Relapse Rate; BID: Twice a Day; CDPS3M: Confirmed Disability Progression Sustained for 3 Months; CDPS6M: Confirmed Disability Progression Sustained for 6 Months; DMF: Dimethyl Fumarate; DMT: Disease-Modifying Therapy; EDSS: Expanded Disability Status Scale; EOD: Every Other Day; GA: Glatiramer Acetate; IFN: Interferon; mcg: Microgram; mg: Milligram; PPR: Proportion of Patients with Relapse; PPR-F: Proportion of Patients Remaining Relapse Free; QD: Once Daily; q4w: Every 4 Weeks; RRMS: Relapsing Remitting Multiple Sclerosis; TIW: Thrice a Week

†N represents randomized population; *Study group not included in our analysis.

Table IV. 2 – Quality assessment results for included RCTs: "risk of bias" summary.

Study	Selection bias		Performance bias	Detection bias	Attrition bias	Reporting bias	Other bias
	Random sequence generation	*Allocation concealment*	*Blinding of participants and personnel*	*Blinding of outcome assessment*	*Incomplete outcome data*	*Selective reporting*	
AFFIRM 2006 **(Polman et al. 2006)**	+	+	+	+	+	+	−
BEYOND 2009 **(O'Connor et al. 2009)**	+	?	−	+	+	+	−
CARE-MS I 2012 **(Cohen et al. 2012)**	+	+	−	?	−	+	−
CARE-MS II 2012 **(Coles et al. 2012)**	+	+	−	?	−	+	−
CONFIRM 2012 **(Fox et al. 2012)**	+	+	−	+	−	−	−
Copolymer I MS Group 1995 **(Johnson et al. 1995)**	?	?	?	?	+	+	−
DEFINE 2012 **(Gold et al. 2012)**	+	+	?	+	−	−	−
FREEDOMS 2010 **(Kappos et al. 2010)**	+	?	?	+	−	+	−
FREEDOMS II 2014 **(Calabresi et al. 2014)**	+	?	+	+	−	+	−
IFNB MS Group 1993 **(Paty & Li 1993)**, **(Paty & Li 2001)**, **([No authors listed] 1993)**	?	?	?	?	?	+	−
MSCRG 1996 **(Jacobs et al. 1996)**	?	?	?	?	+	+	−
PRISMS 1998 **([No authors listed] 1998)**	+	+	?	+	+	+	−
TEMSO 2011 **(O'Connor et al. 2011)**	+	+	+	−	?	−	−

Legend: + Low risk of bias; ? Unclear risk of bias; − High risk of bias.

IV.4.3. FIRST-LINE DMTS FOR RELAPSING-REMITTING MULTIPLE SCLEROSIS (RRMS)

IV.4.3.1. Efficacy: Number Needed to Treat to Benefit (NNTB)

Efficacy results are presented in Figure IV. 2 (and Table IV. 3). All DMTs were more effective than placebo at reducing ARR, excepting IFN-ß-1a-IM. The lowest NNTB was found for IFN-ß-1a-SC and IFN-ß-1b (3; 95% CI 2–4, for both), meaning that three patients need to be treated with IFN-ß-1a SC or IFN-ß-1b rather than placebo to avoid one relapse over 2 years.

With the exception of GA and IFN-ß-1a-IM, the remaining DMTs were associated with higher PPR-F than placebo. Better results were found with IFN-ß-1a-SC (NNTB 7; 95% CI 4–18) and IFN-ß-1b (NNTB 7; 95% CI 3–36), meaning that seven patients need to be treated with these DMT, rather than placebo, for one patient to be free of relapses over 2 years.

Regarding PP-F-CDPS3M, only IFN-ß-1a-SC (NNTB 4; 95% CI 3–7), and teriflunomide (NNTB 15; 95% CI 8–120) were better than placebo. Four patients needed to be treated with IFN-ß-1a SC (or 15 with teriflunomide) rather than placebo to have one additional patient remaining free of disability progression over 2 years. No data was found for IFN-ß-1a-IM and IFN-ß-1b.

Significant differences were not found between GA and IFN-ß-1b on any outcome in the Betaferon Efficacy Yielding Outcomes of a New Dose (BEYOND) study.

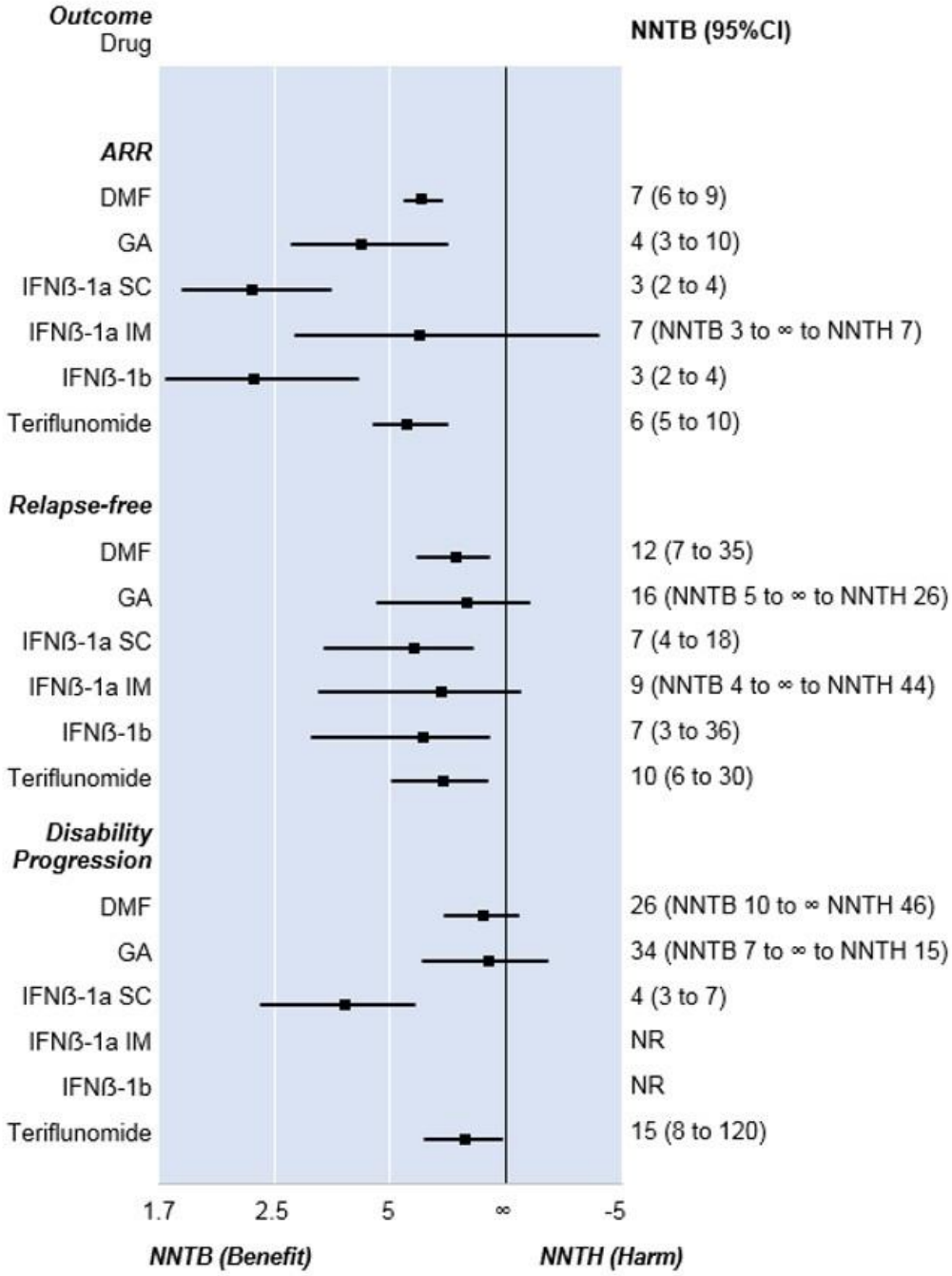


Figure IV. 2 – Numbers needed to treat to benefit (and 95% confidence intervals) for efficacy outcomes with first-line disease-modifying therapies versus comparators.

ARR, annualized relapse rate; CI, confidence interval; DMF, dimethyl fumarate; GA, glatiramer acetate; IFN, interferon; IM, intramuscular; NNTB, number needed to treat to benefit; NNTH, number needed to treat to harm; NR, not reported; SC, subcutaneous.

Table IV. 3 – Data used to estimate NNTB results on efficacy outcomes for first-line disease-modifying therapies.

Comparison	Study	Intervention		Control		RR (95% CI)	Control Event Rate
		Patients, n	Events, n	Patients, n	Events, n		
DMF vs. Placebo							
ARR	DEFINE	659	112	656	236		0.36
	CONFIRM	581	130	599	240		0.40
	Pooled					0.60 (0.51 – 0.69)	0.36[(a)]
PPR-F	DEFINE	410	299	408	220		0.54
	CONFIRM	359	255	363	214		0.59
	Pooled					1.16 (1.05 – 1.28)	0.54[(a)]
PP-F-CDPS3M	DEFINE	410	344	408	298		0.73
	CONFIRM	359	313	363	302		0.83
	Pooled					1.05 (0.97 – 1.14)	0.73[(a)]
GA vs. Placebo							
ARR	Copolymer-1	227	134	232	195	0.70 (0.56 – 0.88)	0.84
PPR-F	Copolymer-1	125	42	126	34	1.25 (0.85 – 1.82)	0.27
PP-F-CDPS3M	Copolymer-1	125	98	126	95	1.04 (0.91 – 1.19)	0.75
IFNß -1a SC vs. Placebo							
ARR	PRISMS	363	318	364	479	0.67 (0.58 – 0.77)	1.32
PPR-F	PRISMS	184	59	187	30	2.00 (1.35 – 2.95)	0.16
PP-F-CDPS3M	PRISMS	184	136	187	86	1.61 (1.35 – 1.92)	0.46
IFNß -1a IM vs. Placebo	MSCRG						
ARR	MSCRG	79	53	78	64	0.82 (0.56 – 1.19)	0.67
PPR-F	MSCRG	85	32	87	23	1.42 (0.91 – 2.22)	0.38
PP-F-CDPS3M	MSCRG	158	NR	143	NR	NR	NR
IFNß -1b vs. Placebo							
ARR	IFNB MS	207	173	209	266	0.66 (0.54 – 0.80)	1.27
PPR-F	IFNB MS	124	36	123	18	1.98 (1.19 – 3.30)	0.15
PP-F-CDPS3M	IFNB MS	124	NR	123	NR	NR	NR
Teriflunomide vs. Placebo							
ARR	TEMSO	628	232	632	341	0.68 (0.58 – 0.81)	0.54
PPR-F	TEMSO	358	217	360	179	1.22 (1.07 – 1.39)	0.50
PP-F-CDPS3M	TEMSO	358	286	360	262	1.10 (1.01 – 1.19)	0.73

ARR, annualized relapse rate; PPR-F, proportion of patients remaining relapse-free; PP-F-CDPS3M, proportion of patients remaining free of confirmed disability progression sustained for 3 months; RR, relative risk.
Number of patient-years of exposure, and number of relapses was used for ARR. Number of randomized patients and number of patients with event was used for PPR-F, and PP-F-CDPS3M.
All Control Event Rates are two-years based. Control Event Rates from a) DEFINE were used for calculating NNTB values for DMF.

IV.4.3.2. Safety: Number Needed to Treat to Harm (NNTH)

Figure IV. 3 (and Table IV. 4) presents safety results. Differences were not found between DMTs and placebo regarding SAEs. The only statistically significant NNTHs for AELD were found for IFN-ß-1b (NNTH 14; 95% CI 2–426) and IFN-ß-1a-SC (NNTH 27; 95% CI 5–57,495). Statistically significant NNTHs were found on several AEs of interest [lowest values with each DMT: DMF, 4 for flushing; GA, 4 for injection-site reaction (ISR); IFN-ß-1a-SC, 2 for ISR; IFN-ß-1a-IM, 5 for influenza-like symptoms; IFN-ß-1b, 3 for ISR; teriflunomide, 11 for alopecia].

In the BEYOND study, the risk of AELD (NNTH 24; 95% CI 13–363) and ISR (NNTH 10; 95% CI 7–22) was higher with GA than with IFN-ß-1b, but the risk of influenza-like symptoms (NNTB 3; 95% CI 3–4) was lower with GA.

IV.4.3.3. Benefit–Risk Ratios: Likelihood to be Helped or Harmed (LHH)

Table IV. 5 summarizes NNTBs, NNTHs, and LHHs for first-line DMTs. Based on NNTH for AELD and NNTB for ARR, the most favourable LHH was found for GA (LHH 59.0), and the least for IFN-ß-1a-IM and IFN-ß-1b (LHH 4.7 for both). However, according to the pool of AEs of interest, IFN-ß-1a-SC was associated with fewer cases of LHH ≤1 (i.e., when risk is higher than benefit).

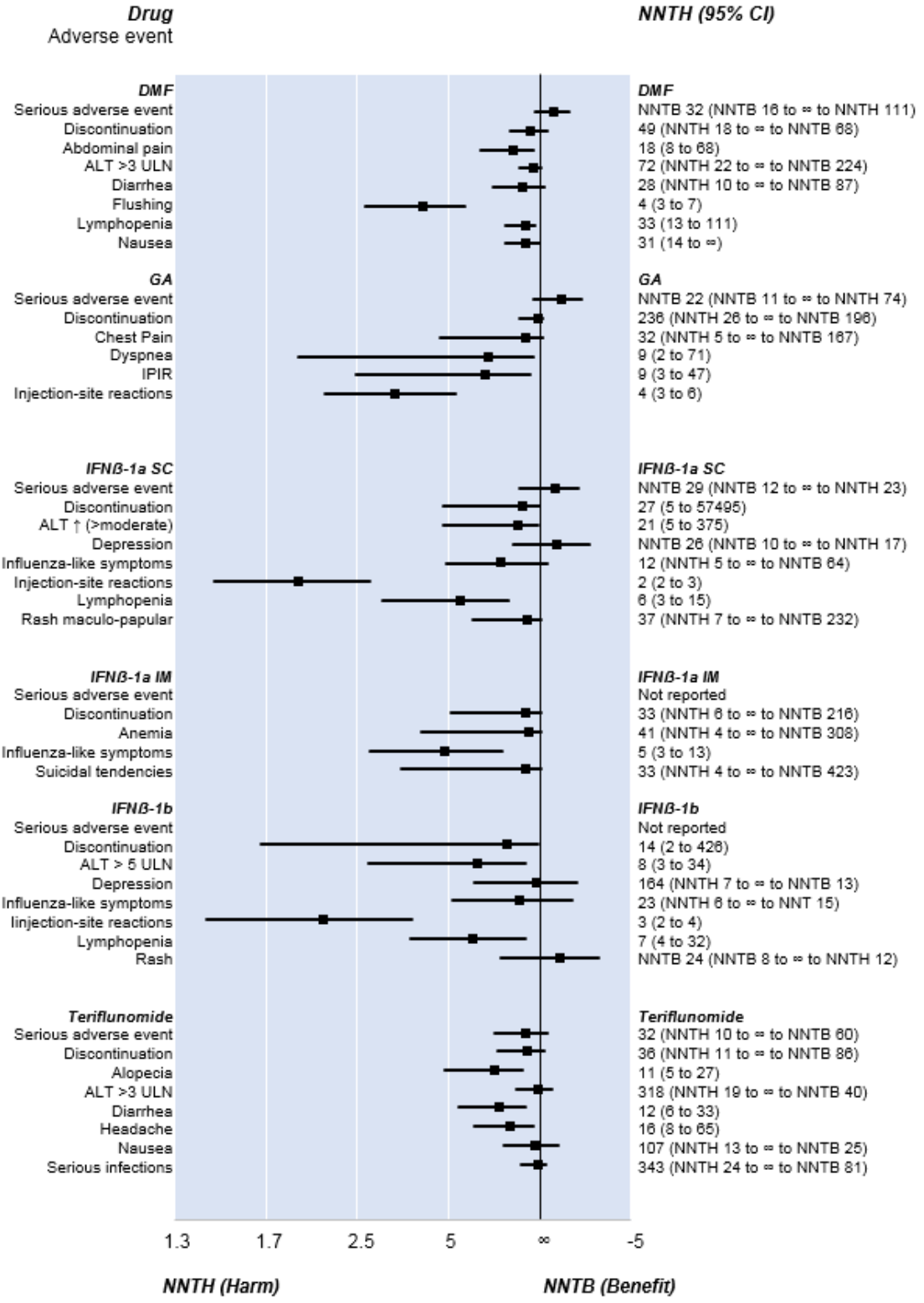


Figure IV. 3 – Numbers needed to treat to harm (and 95% confidence intervals) for safety outcomes with first-line disease-modifying therapies versus comparators.

ALT, alanine transaminase; CI, confidence interval; DMF, dimethyl fumarate; GA, glatiramer acetate; IM, intramuscular; IPIR, immediate post-injection reaction; NNTB, number needed to treat to benefit; NNTH, number needed to treat to harm; NR, not reported; SC, subcutaneous; ULN, upper limit of normal.

Table IV. 4 – Data used to estimate NNTH results on safety outcomes for first-line disease-modifying therapies.

Comparison	Study	Intervention		Control		RR (95% CI)	Control Event Rate
		Randomized patients, n	Patients with event, n	Randomized patients, n	Patients with event, n		
DMF vs. Placebo							
SAE	DEFINE	410	74	408	86		
	CONFIRM	359	61	363	79		
	Pooled					0.85 (0.69 - 1.04)	0.21[(a)]
AELD	DEFINE	410	65	408	55		
	CONFIRM	359	44	363	38		
	Pooled					1.15 (0.89 - 1.49)	0.13[(a)]
Abdominal pain	DEFINE	410	46	408	22		
	CONFIRM	359	NR	363	NR		
	Pooled					2.08 (1.28 - 3.39)	0.05[(a)]
ALT >3 ULN	DEFINE	410	25	408	12		
	CONFIRM	359	23	363	20		
	Pooled					1.48 (0.85 - 2.57)	0.03[(a)]
Diarrhoea	DEFINE	410	62	408	55		
	CONFIRM	359	45	363	28		
	Pooled					1.27 (0.91 - 1.76)	0.13[(a)]
Flushing	DEFINE	410	154	408	20		
	CONFIRM	359	110	363	13		
	Pooled					6.20 (4.37 - 8.79)	0.05[(a)]
Lymphopenia (grade ≥3)	DEFINE	410	16	408	4		
	CONFIRM	359	18	363	4		
	Pooled					4.12 (1.92 - 8.85)	0.01[(a)]
	DEFINE	410	53	408	38		
	CONFIRM	359	40	363	29		
	Pooled					1.35 (1.00 - 1.82)	0.09[(a)]
GA vs. Placebo							
SAE	CONFIRM	351	60	363	79	0.79 (0.58 – 1.06)	0.22
AELD*	Copolymer-1	125	5	126	1		
	CONFIRM	351	35	363	38		
	Pooled	476	40	489	39	1.53 (0.36 – 6.59)	0.01[(b)]
Chest pain (without flushing)	Copolymer-1	125	6	126	2	3.02 (0.62 - 14.70)	0.02
Dyspnoea	Copolymer-1	125	16	126	2	8.06 (1.89 - 34.34)	0.02
Immediate-post injection reaction	Copolymer-1	125	19	126	4	4.79 (1.68 - 13.68)	0.03
Injection-site reaction	Copolymer-1	125	113	126	74	1.54 (1.32 - 1.80)	0.59
IFNß -1a SC vs. Placebo							
SAE	PRISMS	184	22	187	29	0.77 (0.46 -1.29)	0,16
AELD	PRISMS	184	9	187	2	4.57 (1.00 - 20.88)	0,01
ALT increase (moderate to severe)	PRISMS	184	12	187	3	4.07 (1.17 - 14.17)	0,02
Depression	PRISMS	184	44	187	52	0.86 (0.61 - 1.22)	0,28
Influenza-like symptoms	PRISMS	184	109	187	95	1.17 (0.97 - 1.40)	0,51
Injection-site reaction	PRISMS	184	169	187	73	2.35 (1.96 - 2.83)	0,39
Lymphopenia	PRISMS	184	53	187	21	2.56 (1.62 - 4.07)	0,11
Rash maculopapular	PRISMS	184	8	187	3	2.71 (0.73 -10.06)	0,02
IFNß -1a IM vs. Placebo							
SAE	MSCRG	158	NR	143	NR	NR	NR
AELD	MSCRG	158	7	143	2	3.17 (0.67-15.00)	0.01
Anaemia	MSCRG	158	5	143	1	4.53 (0.54 - 38.28)	0.01
Influenza-like symptoms	MSCRG	158	96	143	57	1.52 (1.20 - 1.93)	0.40
Suicidal tendencies	MSCRG	158	6	143	1	5.43 (0.66 - 44.56)	0.01
IFNß -1b vs. Placebo							
SAE	IFNB MS Study	124	NR	123	NR	NR	NR
AELD	IFNB MS Study	124	10	123	1	9.92 (1.29 - 76.32)	0.01
ALT >5 ULN	IFNB MS Study	124	24	123	7	3.40 (1.52 - 7.60)	0.06
Depression	IFNB MS Study	124	31	123	30	1.03 (0.66 - 1.58)	0.24
Influenza-like	IFNB MS Study	124	65	123	59	1.09 (0.85 - 1.40)	0.48

Comparison	Study	Intervention		Control		RR (95% CI)	Control Event Rate
		Randomized patients, n	Patients with event, n	Randomized patients, n	Patients with event, n		
symptoms							
Injection-site reaction	IFNB MS Study	124	96	123	37	2.57 (1.93 - 3.42)	0.30
Lymphopenia	IFNB MS Study	124	99	123	80	1.23 (1.05 - 1.44)	0.65
Rash	IFNB MS Study	124	34	123	39	0.86 (0.59 - 1.27)	0.32
Teriflunomide vs. Placebo							
SAE	TEMSO	358	57	360	46	1.25 (0.87 – 1.79)	0.13
AELD	TEMSO	358	39	360	29	1.35 (0.86 – 2.14)	0.08
Alopecia	TEMSO	358	47	360	12	3.94 (2.13 – 7.30)	0.03
ALT >3 ULN	TEMSO	358	24	360	23	1.05 (0.60 – 1.82)	0.06
Diarrhoea	TEMSO	358	64	360	32	2.01 (1.35 – 3.00)	0.09
Nausea	TEMSO	358	49	360	26	1.90 (1.21 – 2.98)	0.07
Headache	TEMSO	358	67	360	64	1.05 (0.77 – 1.43)	0.18
Serious infections	TEMSO	358	9	360	8	1.14 (0.44 – 2.90)	0.02

AELD, adverse event leading to discontinuation of the study drug; RR, relative risk; SAE, serious adverse event; ULN, upper limit of normal.
All Control Event Rates are two-years based. Control Event Rate from a) DEFINE; b) Copolymer 1 Trial.
*Data available from Copolymer 1 Trial and CONFIRM Study.
Number of randomized patients and number of patients with event was used.

Table IV. 5 – NNTHs, NNTBs and LHHs for first-line disease-modifying therapies versus comparators on outcomes of safety and efficacy.

		Annualized Relapse Rate	Relapse-free	Free of disability progression	
FIRST-LINE DMTs					
	NNTH	**LHH**	**LHH**	**LHH**	
Dimethyl fumarate		(NNTB = 7)	(NNTB = 12)	(NNTB = 26),a	
Serious adverse event	-32	NA	NA	NA	b
Discontinuation	49	7,0	4,1	1,9	b
Abdominal pain	18	2,6	1,5	0,7	c
ALT >3 ULN	72	10,3	6,0	2,8	b
Diarrhoea	28	4,0	2,3	1,1	b
Flushing	4	0,6	0,3	0,2	c
Lymphopenia	33	4,7	2,8	1,3	
Nausea	31	4,4	2,6	1,2	
Glatiramer acetate		(NNTB = 4)	(NNTB = 16),a	(NNTB = 34),a	
Serious adverse event	-22	NA	NA	NA	b
Discontinuation	236	59,0	14,8	6,9	b
Chest pain (without flushing)	32	8,0	2,0	0,9	b,c
Dyspnoea	9	2,3	0,6	0,3	c
Immediate-post injection reaction	9	2,3	0,6	0,3	c
Injection-site reaction	4	1,0	0,3	0,1	c
IFN-β -1a SC		(NNTB = 3)	(NNTB = 7)	(NNTB = 4)	
Serious adverse event	-29	NA	NA	NA	b
Discontinuation	27	9,0	3,9	6,8	
ALT increase (moderate to severe)	21	7,0	3,0	5,3	
Depression	-26	NA	NA	NA	b
Influenza-like symptoms	12	4,0	1,7	3,0	b
Injection-site reaction	2	0,7	0,3	0,5	c
Lymphopenia	6	2,0	0,9	1,5	c
Rash maculopapular	37	12,3	5,3	9,3	b
IFN-β -1a IM		(NNTB = 7),a	(NNTB = 9),a	(NNTB = NA)	
Serious adverse event	NR	NA	NA	NA	
Discontinuation	33	4,7	3,7	NA	b
Anaemia	41	5,9	4,6	NA	b
Influenza-like symptoms	5	0,7	0,6	NA	c
Suicidal tendencies	33	4,7	3,7	NA	b
IFN-β -1b		(NNTB = 3)	(NNTB = 7)	(NNTB = NA)	
Serious adverse event	NR	NA	NA	NA	
Discontinuation	14	4,7	2,0	NA	
ALT >5 ULN	8	2,7	1,1	NA	
Depression	164	54,7	23,4	NA	b
Influenza-like symptoms	23	7,7	3,3	NA	b
Injection-site reaction	3	1,0	0,4	NA	c
Lymphopenia	7	2,3	1,0	NA	c
Rash	-24	NA	NA	NA	b
Teriflunomide		(NNTB = 6)	(NNTB = 10)	(NNTB = 15)	
Serious adverse event	32	5,3	3,2	2,1	b
Discontinuation	36	6,0	3,6	2,4	b
Alopecia	11	1,8	1,1	0,7	c
ALT >3 ULN	318	53,0	31,8	21,2	b
Diarrhoea	12	2,0	1,2	0,8	c
Headache	16	2,7	1,6	1,1	
Nausea	107	17,8	10,7	7,1	b
Serious infections	343	57,2	34,3	22,9	b

ALT, alanine aminotransferase; DMT, disease-modifying therapy; IFN, interferon; IM, intramuscular; LHH, likelihood to be helped or harm; NA, not applicable; NNTB, number needed to treat to benefit; NNTH, number needed to treat to harm; NR, not reported; SC, subcutaneous; ULN, upper limit of normal.

a) NNTB not statistically significant.

b) NNTH not statistically significant.

c) Benefit ≤ Risk, for at least one outcome of efficacy.

IV.4.4. SECOND-LINE DMTS AND HIGHLY ACTIVE RRMS

IV.4.4.1. Efficacy: NNTB

Results are presented in Figure IV. 4 (and Table IV. 6). All DMTs were better than comparators on ARR and PPR-F, with the lowest NNTBs reported for natalizumab (NNTB 2; 95% CI 2–3; and NNTB 4; 95% CI 3–6). Data was not available for alemtuzumab on PP-F-CDPS3M. The NNTB versus placebo for PP-F-CDPS3M was estimated at 9 (95% CI 6–19) with natalizumab.

IV.4.4.2. Safety: NNTH

The risk of AELD was lower with alemtuzumab than with its comparator (NNTB 22; 95% CI 17–41), while no significant differences were found between fingolimod or natalizumab and placebo (Figure IV. 5; Table IV. 7). Statistically significant NNTHs were found on AEs of interest versus comparators (lowest values with each DMT: alemtuzumab, 6 for IAR; fingolimod, 22 for abnormal liver function; natalizumab, 20 for lymphocytosis).

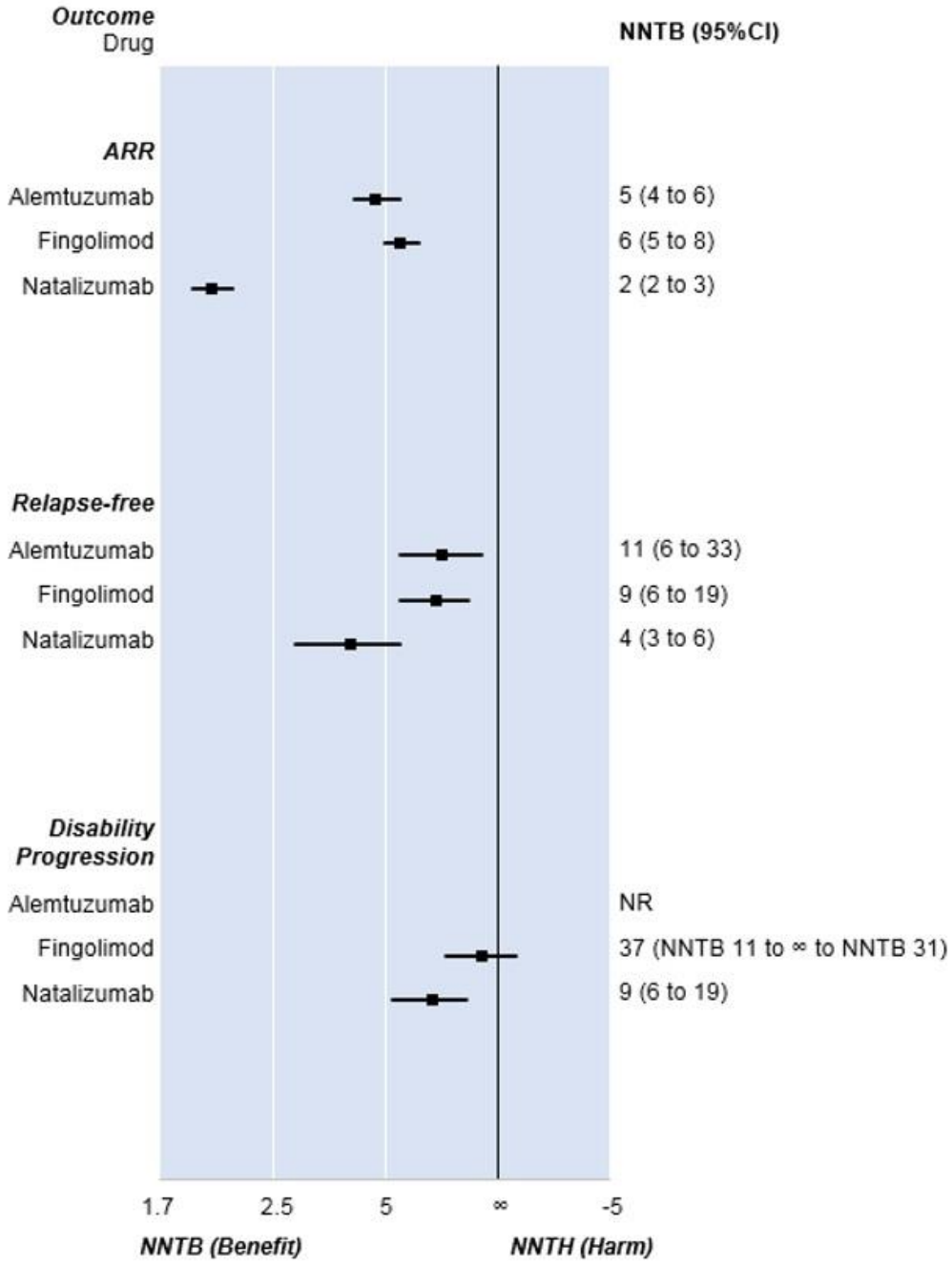


Figure IV. 4 – Numbers needed to treat to benefit (and 95% confidence intervals) for efficacy outcomes with second-line or highly-active RRMS disease-modifying therapies versus comparators.

ARR, annualized relapse rate; CI, confidence interval; NNTB, number needed to treat to benefit; NNTH, number needed to treat to harm; NR, not reported.

Table IV. 6 – Data used to estimate NNTB results on efficacy outcomes for second-line or highly-active RRMS disease-modifying therapies.

Comparison	Study	Intervention		Control		RR (95% CI)	Control Event Rate
		Patients, n	Events, n	Patients, n	Events, n		
Alemtuzumab vs. IFNß -1a SC							
ARR	CARE-MS I	661	119	313	122		0.39
	CARE-MS II	908	236	387	201		0.52
	Pooled					0.58 (0.51 – 0.66)	0.52[(a)]
PPR-F	CARE-MS I	376	292	187	110		0.59
	CARE-MS II	435	285	202	94		0.47
	Pooled					1.21 (1.07 – 1.37)	0.47[(a)]
PP-F-CDPS3M	CARE-MS I	376	NR	187	NR		
	CARE-MS II	435	NR	202	NR		
	Pooled					NR	NR
Fingolimod vs. Placebo							
ARR	FREEDOMS	794	143	750	300		0.40
	FREEDOMS II	630	132	610	244		0.40
	Pooled					0.57 (0.50 – 0.65)	0.40[(b)]
PPR-F	FREEDOMS	425	299	418	191		0.46
	FREEDOMS II	358	256	355	187		0.53
	Pooled					1.21 (1.10 – 1.33)	0.53[(b)]
PP-F-CDPS3M	FREEDOMS	425	350	418	317		0.76
	FREEDOMS II	358	267	355	252		0.71
	Pooled					1.04 (0.95 – 1.13)	0.71[(b)]
Natalizumab vs. Placebo							
ARR	AFFIRM	1202	289	596	447	0.32 (0.28 – 0.37)	0.24
PPR-F	AFFIRM	627	454	315	146	1.56 (1.37 – 1.78)	0.72
PP-F-CDPS3M	AFFIRM	627	520	315	224	1.17 (1.08 – 1.26)	0.71

ARR, annualized relapse rate; PPR-F, proportion of patients remaining relapse-free; PP-F-CDPS3M, proportion of patients remaining free of confirmed disability progression sustained for 3 months; RR, relative risk.

Number of patient-years of exposure, and number of relapses was used for ARR. Number of randomized patients and number of patients with event was used for PPR-F, and PP-F-CDPS3M.

All Control Event Rates are two-years based. Control Event Rates from a) CARE-MS II, and b) FREEDOMS II were used for calculating NNTB values for alemtuzumab, and fingolimod, respectively.

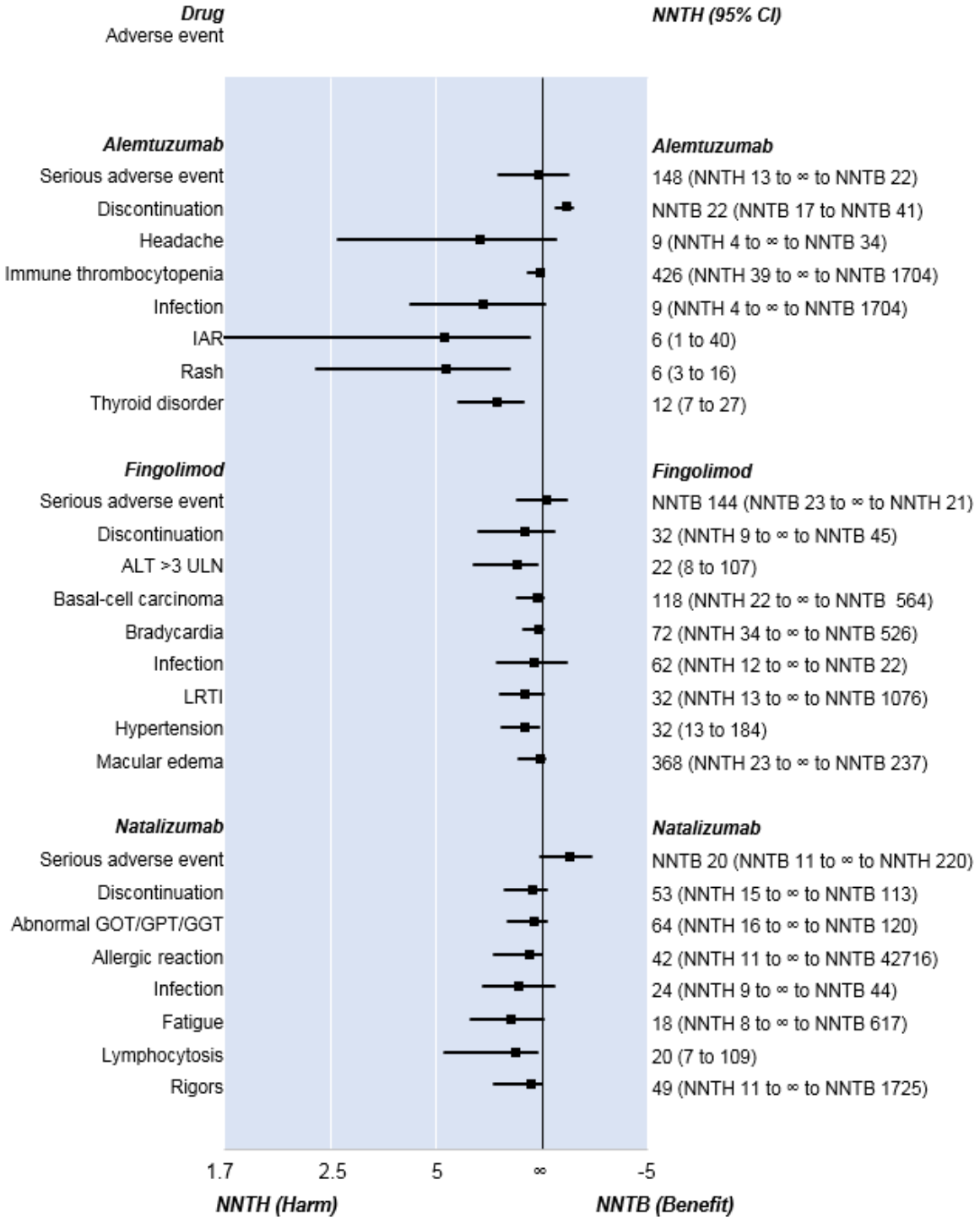


Figure IV. 5 – Numbers needed to treat to harm (and 95% confidence intervals) for safety outcomes with second-line or highly-active RRMS disease-modifying therapies versus comparators.

CI, confidence interval; GGT, gamma-glutamyl transferase; GOT, glutamic oxaloacetic transaminase; GPT, glutamate-pyruvate transaminase; IAR, infusion-associated reaction; LRTI, low respiratory tract infection; NNTB, number needed to treat to benefit; NNTH, number needed to treat to harm; NR, not reported; ULN, upper limit of normal.

Table IV. 7 – Data used to estimate NNTH results on safety outcomes for second-line or highly-active RRMS disease-modifying therapies.

Comparison	Study	Intervention		Control		RR (95% CI)	Control Event Rate
		Randomized patients, n	Patients with event, n	Randomized patients, n	Patients with event, n		
Alemtuzumab vs. IFNß -1a SC							
SAE	CARE-MS I	376	69	187	27		
	CARE-MS II	435	85	202	44		
	Pooled					1.03 (0.78 - 1.37)	0.22[a]
Discontinuation	CARE-MS I	376	5	187	11		
	CARE-MS II	435	14	202	15		
	Pooled					0.37 (0.20 - 0.66)	0.07[a]
Headache	CARE-MS I	376	86	187	35		
	CARE-MS II	435	230	202	36		
	Pooled					1.65 (0.86 - 3.17)	0.18[a]
Immune thrombocytopenia	CARE-MS I	376	3	187	0		
	CARE-MS II	435	3	202	0		
	Pooled					3.35 (0.41 - 27.12)	0.001[b]
Infection	CARE-MS I	376	253	187	85		
	CARE-MS II	435	334	202	134		
	Pooled					1.17 (0.99 - 1.38)	0.66[a]
Infusion-associated reaction	CARE-MS I	376	338	187	0		
	CARE-MS II	435	393	202	0		
	Pooled					185.22 (26.14 - 1312.49)	0.001[b]
Rash	CARE-MS I	376	44	187	7		
	CARE-MS II	435	193	202	11		
	Pooled					4.36 (2.15 - 8.81)	0.05[a]
Thyroid disorder	CARE-MS I	376	68	187	12		
	CARE-MS II	435	69	202	10		
	Pooled					2.70 (1.75 - 4.17)	0.05[a]
Fingolimod vs. Placebo							
SAE	FREEDOMS	425	43	418	56		
	FREEDOMS II	358	53	355	45		
	Pooled					0.95 (0.65 - 1.38)	0.13[c]
AELD	FREEDOMS	425	32	418	32		
	FREEDOMS II	358	66	355	37		
	Pooled					1.30 (0.79 - 2.15)	0.10[c]
ALT >3 ULN	FREEDOMS	425	NR	418	NR		
	FREEDOMS II	358	25	355	8		
	Pooled					3.10 (1.42 - 6.78)	0.02[e]
Basal-cell carcinoma	FREEDOMS	425	4	418	3		
	FREEDOMS II	358	10	355	2		
	Pooled					2.50 (0.69 - 9.15)	0.01[c]
Bradycardia	FREEDOMS	425	9	418	3		
	FREEDOMS II	358	3	355	3		
	Pooled					1.88 (0.67 - 5.31)	0.01[c]
Infection	FREEDOMS	425	NR	418	NR		
	FREEDOMS II	358	263	355	255		
	Pooled					1.02 (0.93 - 1.12)	0.72[c]
Low respiratory tract infection	FREEDOMS	425	41	418	25		
	FREEDOMS II	358	38	355	30		
	Pooled					1.38 (0.99 - 1.95)	0.08[c]
Hypertension	FREEDOMS	425	26	418	16		
	FREEDOMS II	358	32	355	11		
	Pooled					2.03 (1.18 - 3.50)	0.03[c]
Macular oedema	FREEDOMS	425	0	418	0		
	FREEDOMS II	358	3	355	2		
	Pooled					1.48 (0.25 - 8.82)	0.01[c]
Natalizumab vs. Placebo							
SAE	AFFIRM	627	119	312	75	0.79 (0.61 - 1.02)	0.24
AELD	AFFIRM	627	38	312	13	1.45 (0.79 - 2.69)	0.04
Abnormal GOT/GPT/GGT	AFFIRM	627	30	312	10	1.49 (0.74 - 3.01)	0.03
Allergic reaction	AFFIRM	627	23	312	4	2.86 (0.99 - 8.20)	0.01
Infection	AFFIRM	627	424	312	198	1.07 (0.96 - 1.18)	0.63

Comparison	Study	Intervention		Control		RR (95% CI)	Control Event Rate
		Randomized patients, n	Patients with event, n	Randomized patients, n	Patients with event, n		
Fatigue	AFFIRM	627	169	312	66	1.27 (0.99 - 1.64)	0.21
Lymphocytosis	AFFIRM	627	38	312	3	6.30 (1.96 - 20.26)	0.01
Rigors	AFFIRM	627	19	312	3	3.15 (0.94 - 10.57)	0.01

AELD, adverse event leading to discontinuation of the study drug; GGT, gamma-glutamyl transferase; GOT, glutamic oxaloacetic transaminase; GPT, glutamate-pyruvate transaminase; RR, relative risk; SAE, serious adverse event; ULN, upper limit of normal.
All Control Event Rates are two-years based. Control Event Rate from a) CARE-MS II; b) control event rate of 0.1% was assumed to allow calculating NNTH; c) FREEDOMS II.
Number of randomized patients and number of patients with event was used.

IV.4.4.3 Benefit–Risk Ratios: LHH

The results for second-line and highly active RRMS are presented in Table IV. 8. Based on NNTB results for ARR and NNTH results for AELD, LHHs were estimated at 53 for natalizumab, and 32 for fingolimod. LHH could not be calculated on this safety outcome for alemtuzumab. LHH was >1 in all comparisons carried out for natalizumab.

Table IV. 8 – NNTHs, NNTBs and LHHs for second-line or highly-active RRMS disease-modifying therapies versus comparators on outcomes of safety and efficacy.

		Annualized Relapse Rate	Relapse-free	Free of disability progression	
SECOND-LINE DMTs					
	NNTH	LHH	LHH	LHH	
Alemtuzumab		(NNTB = 5)	(NNTB = 11)	(NNTB = NA)	
Serious adverse event	148	29,6	13,5	NA	b
Discontinuation	-22	NA	NA	NA	b
Headache	9	1,8	0,8	NA	b,c
Immune thrombocytopenia	426	85,2	38,7	NA	b
Infection	9	1,8	0,8	NA	b,c
Infusion-associated reaction	6	1,2	0,5	NA	c
Rash	6	1,2	0,5	NA	c
Thyroid disorder	12	2,4	1,1	NA	
Fingolimod		(NNTB = 6)	(NNTB = 9)	(NNTB = 37),a	
Serious adverse event	-144	NA	NA	NA	b
Discontinuation	32	5,3	3,6	0,9	b,c
ALT >3 ULN	22	3,7	2,4	0,6	c
Basal-cell carcinoma	118	19,7	13,1	3,2	b
Bradycardia	72	12,0	8,0	1,9	b
Infection	62	10,3	6,9	1,7	b
Low respiratory tract infection	32	5,3	3,6	0,9	b,c
Hypertension	32	5,3	3,6	0,9	c
Macular oedema	368	61,3	40,9	9,9	b
Natalizumab		(NNTB = 2)	(NNTB = 4)	(NNTB = 9)	
Serious adverse event	-20	NA	NA	NA	b
Discontinuation	53	26,5	13,3	5,9	b
Abnormal GOT/GPT/GGT	64	32,0	16,0	7,1	b
Allergic reaction	42	21,0	10,5	4,7	b
Infection	24	12,0	6,0	2,7	b
Fatigue	18	9,0	4,5	2,0	b
Lymphocytosis	20	10,0	5,0	2,2	
Rigors	49	24,5	12,3	5,4	b

ALT, alanine aminotransferase; DMT, disease-modifying therapy; GGT, gamma-glutamyl transferase; GOT, glutamic oxaloacetic transaminase; GPT, glutamate-pyruvate transaminase; LHH, likelihood to be helped or harmed; NA, not applicable; NNTB, number needed to treat to benefit; NNTH, number needed to treat to harm; RRMS, relapsing–remitting multiple sclerosis; ULN, upper limit of normal.
a) NNTB not statistically significant.
b) NNTH not statistically significant.
c) Benefit ≤ Risk, for at least one outcome of efficacy.

IV.5. DISCUSSION

Several systematic reviews and meta-analyses have been published discussing efficacy and safety profiles of DMTs in RRMS. However, their results are mainly expressed as relative measures of effect (Tramacere et al. 2015), (Hadjigeorgiou et al. 2013), (Hutchinson et al. 2014), (Roskell et al. 2012). The major problem with relative effect measures is that they do not reflect baseline risks (i.e., without intervention), making it impracticable to discriminate large from small treatment effects, and leading sometimes to misleading conclusions (Klawiter, Cross & Naismith 2009), (Citrome 2010).

The most recently published studies have been reporting higher relative risk reductions (RRR) with DMTs versus placebo on outcomes of efficacy in RRMS patients. For example, RRR on ARR versus placebo was estimated at 33% with IFN-ß-1a-SC (O'Connor et al. 2011), and 44–53% with DMF (Fox et al. 2012), (Gold et al. 2012). Such results may give the perception that DMF is more effective than IFN-ß-1a-SC. However, the results of the present study suggest that the number of patients needed to treat with DMF to avoid one relapse over 2 years (NNTB 7) is >2-fold the number of patients one would need to treat with IFN-ß-1a-SC (NNTB 3). This example illustrates the usefulness of absolute measures of effect.

Lower values of NNTB indicate better outcomes. Across first-line DMTs, IFN-ß-1a-SC was consistently associated with the lowest values of NNTB for all outcomes of efficacy. GA failed to demonstrate significant benefits on PPRF and PP-F-CDPS3M, as did IFN-ß-1a-IM on both ARR and PPR-F, and DMF on PP-F-CDPS3M. Compared with placebo, previous studies have found lower values of NNTB with IFN than with GA (Francis 2004). Nevertheless, mixed treatment comparisons suggested that DMF was more effective than IFNs and GA, with IFNs and GA having similar efficacy in terms of relapse reduction (Hutchinson et al. 2014), (Roskell et al. 2012).

For NNTH, higher values are better. The use of first-line DMTs did not increase the risk of SAEs, and only IFN-ß-1b (NNTH 14) and IFN-ß-1a-SC (NNTH 27) significantly increased the risk of AELD. Among injectable DMTs, a slightly lower NNTH was found for injection-site reactions with IFNs than with GA; and only IFN-ß-1a-IM increased influenza-like symptoms (NNTH 5). Regarding oral formulations, flushing with DMF (NNTH 4), and alopecia (NNTH 11) with teriflunomide, were the most common AEs.

In order to assess benefit–risk ratios, NNTBs and NNTHs were compared. Negative benefit–risk ratios, i.e. LHH ≤1, were less often encountered for teriflunomide (only for alopecia and diarrhoea against PP-F-CDPS3M) than with other first-line DMTs. The first-line DMT with less favourable benefit–risk ratios seems to be GA. Regarding DMF, LHH ≤1 was found only for flushing versus any efficacy outcome, and abdominal pain versus PPR-F. Injection-site reactions and lymphopenia were the only AEs leading to LHH values ≤1 in the case of IFNs. However, if the cut-off LHH was ≥2 (i.e., the expected number of patients that will benefit from treatment is at least twice the number of patients that will be harmed by it), IFN-ß-1a-SC seems to have the most favourable benefit-risk ratio.

Among second-line DMTs, natalizumab was consistently associated with lower NNTB values. However, alemtuzumab was compared with IFN-ß-1a-SC instead of placebo, which may have contributed to higher NNTBs with alemtuzumab. Thus, caution is needed when interpreting these results. A network meta-analysis suggested that alemtuzumab is the most effective DMT in reducing the recurrence of relapses, as compared with placebo (Tramacere et al. 2015).

Significant NNTHs were not found for SAEs in a second-line setting. Regarding AELDs, alemtuzumab had a protective effect versus IFN-ß-1a-SC (NNTB 22). Overall, the lowest NNTHs were found for alemtuzumab, particularly due to infusion-associated reactions and rash. As compared with placebo, significant NNTHs were found only for increased alanine transaminase and hypertension with fingolimod, and lymphocytosis with natalizumab.

According to LHH results, natalizumab seems to have the most favourable benefit–risk ratio, with LHHs >1 irrespective of the outcomes analysed. Fingolimod was associated with LHHs <1 for several safety outcomes (AELD, increased alanine transaminase, low respiratory tract infection, and hypertension) against PP-F-CDPS3M. The less favourable benefit–risk ratios were found for alemtuzumab, with LHH values <1 when headache, infection, infusion-associated reaction, and rash were weighed against PPR-F.

A few limitations and some considerations should be taken into account when interpreting the results. DMTs were classified as first- and second-line/highly active options according to the therapeutic indications approved in Europe. However, the EMA and the FDA differ with regards to the recommendations of use for fingolimod and natalizumab. Despite being reserved for patients with highly active disease (after having received other DMTs) or with rapidly evolving RRMS in Europe (GILENYA 2016a), (TYSABRI 2016), both

fingolimod and natalizumab can be used as first-line treatments in the USA (GILENYA 2016b), (TYSABRI 2012). This issue may limit the applicability of the results to settings using European recommendations.

Disability progression was measured differently among the studies. For example, disability progression was defined as an increase of at least 1.0 point on the EDSS in patients with a baseline score of 1.0 or higher in DMF studies, but an increase of 0.5 points in patients with a baseline EDSS score >5.5 would be sufficient to assign the same outcome in teriflunomide studies (Fox et al. 2012), (Gold et al. 2012), (O'Connor et al. 2011).

Further, the characteristics of patients treated during clinical practice are often different from those included in RCTs. Thus, NNT values are likely to be different in clinical practice (Francis 2004). Moreover, clinicians must be aware that the NNT is always dependent on a control event rate. For example, drug A and drug B were evaluated in two placebo-controlled clinical trials (Study A and Study B, respectively), both with a follow-up of 1 year. The ARR for patients receiving drug A and drug B was the same (ARR 0.25), but the ARR for patients receiving placebo was estimated at 0.5 in Study A and 0.75 in Study B. Despite both drugs having produced the same ARR, the NNTB versus placebo over 1 year of treatment would be more favourable with drug B (NNTB 2) than with drug A (NNTB 4) due to a worse placebo performance in Study B. Thus, it is of utmost importance to acknowledge the effect of control event rates when interpreting NNT estimates.

According to the results of the present study, DMF (ARR 0.17–0.22) would be apparently more advantageous than IFN-ß-1a-SC (ARR 0.88), but the resulting NNTB versus placebo in each RCT tells the opposite (NNTB 7 for DMF and 3 for IFN-ß-1a-SC). Indeed, subjects have a higher placebo response and therefore a lower ARR in more recent trials than in earlier ones (Klawiter, Cross & Naismith 2009). Using these examples, the ARR of placebo was estimated at 0.36–0.40 in DMF trials, and at 1.32 in the IFN-ß-1a-SC trial (Fox et al. 2012), (Gold et al. 2012), ([No authors listed] 1998). A growth of the placebo effect has also been seen in other settings, such as in RCTs conducted with neuropathic pain drugs in the USA (Tuttle et al. 2015). An issue that might have been contributing to the apparently enhanced efficacy of DMTs and placebo in more recent RCTs is the inclusion of patients with less severe disease and/or earlier phases of RRMS, whom are given more timely diagnosis, particularly since the widespread use of MRI (Caucheteux et al. 2015).

The safety analysis was challenging, and some considerations need to be pointed out. Only two parameters (SAE and AELD) were common to all DMTs, with only AELD having data available for all drugs. One issue that needs to be taken into account when considering

data on AELD from RCTs is that some of those events would not necessarily lead to treatment discontinuation in daily clinical practice. For example, in the active-controlled teriflunomide studies, any patient receiving IFN with an ALT increase >3 x the upper limit of normal (ULN) was required to discontinue treatment (Vermersch et al. 2014). However, in clinical practice, some patients on IFN presenting liver enzyme elevations would have continued on the medication, as most of these laboratory abnormalities resolve spontaneously or after dose reduction (Oh & O'Connor 2014). Further, caution is needed when comparing results between DMTs owing to heterogeneous definitions of AEs. For example, hepatotoxicity was not assessed in the same way for all DMTs, with reports of ALT increased >3 x ULN with DMF and >5 x ULN with IFN-ß-1b. In addition, in spite of being traditionally considered as the highest quality evidence, RCTs have limitations in capturing safety information, particularly rare and/or long-term latency AEs (Rawlins 2008), (Hammad et al. 2013-b) (Hammad et al. 2013). Thus, other safety problems that were not reported in selected RCTs deserve further investigation, namely progressive multifocal leukoencephalopathy (PML).

After decades of use, no increase in risk of PML was linked with IFNs or GA (Sheremata, Brown & Rammohan 2015). However, natalizumab, fingolimod, and DMF have been associated with cases of PML in MS patients (Bloomgren et al. 2012), (EMA 2015c), (EMA 2015d). The risk of PML can be stratified according to some factors. Prolonged lymphopenia and anti-JCV (John Cunningham virus) antibody index >0.9 appear to influence PML development (Faulkner 2015), (McGuigan et al. 2016). Previous immunosuppression and treatment duration >2 years are important risk factors for natalizumab (Bloomgren et al. 2012). Regarding DMF, EMA published guidelines recommending regular lymphocyte counts, after noticing three cases of PML in patients presenting counts <500/μL (EMA 2015d). However, a case with counts between 500 and 800/μL was later reported (Nieuwkamp et al. 2015). In the case of fingolimod, patient stratification according to lymphocyte counts is not reasonable given its mode of action (Ingwersen, Aktas & Hartung 2016).

The incidence of PML with DMTs can be estimated based on observational data. NNTH values can be computed by applying PML incidence in the general population (0.3/100,000 person-years) (Arkema, van Vollenhoven & Askling 2012). Using overall incidence of PML in natalizumab patients (101/100,000 person-years) (Bloomgren et al. 2012), the NNTH would be 990 (95% CI 318–2354).

In April 2016, EudraVigilance (a database of all suspected unexpected serious adverse reactions) contained 33 cases of PML reported for fingolimod. Given the most up to-date

public numbers, fingolimod was used for at least 241,300 person-years (Gilenyaworldwatch.com 2016). According to the manufacturer of DMF, four cases of PML were reported for 220,000 person-years of exposure. Thus, incidence rates of PML are estimated at 13.7 and 1.8 cases per 100,000 person-years for fingolimod and DMF, respectively. NNTH values would be 7459 (95% CI 2242–19,856) for fingolimod, and 64,551 (95% CI 12,228–1,089,731) for DMF.

In clinical practice, evaluating risk factors for developing PML is determinant for making treatment decisions. JCV status is particularly relevant, limiting the use of natalizumab in case of a positive or high-index result (Ingwersen, Aktas & Hartung 2016), (Plavina et al. 2014). Thus, the results of the present study are primarily applicable to patients without risk factors for PML or other contraindications.

This study shows that NNTB, NNTH, and LHH are valuable tools for use in benefit–risk assessments. These metrics have the advantage of reflecting baseline risks of events into clinically useful results, which can be immediately perceived by clinicians. In conclusion, the overall results suggest that, as compared with placebo, IFN-ß-1aSC has the most favourable benefit–risk ratio among first-line treatment options for RRMS. Natalizumab was associated with better benefit–risk ratios than the other DMTs approved in second-line or in highly active RRMS. Continuous research needs to be carried out upon the production of new and/or updated evidence on efficacy and safety of DMTs.

IV.6. REFERENCES

[NO AUTHORS LISTED], 1993. INTERFERON BETA-1B IS EFFECTIVE IN RELAPSING-REMITTING MULTIPLE SCLEROSIS. I. CLINICAL RESULTS OF A MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED TRIAL. THE IFNB MULTIPLE SCLEROSIS STUDY GROUP. NEUROLOGY, 43(4), PP. 655-661.

[NO AUTHORS LISTED], 1998. RANDOMISED DOUBLE-BLIND PLACEBO-CONTROLLED STUDY OF INTERFERON BETA-1A IN RELAPSING/REMITTING MULTIPLE SCLEROSIS. PRISMS (PREVENTION OF RELAPSES AND DISABILITY BY INTERFERON BETA-1A SUBCUTANEOUSLY IN MULTIPLE SCLEROSIS) STUDY GROUP. LANCET, 352(9139), PP. 1498-1504.

ALTMAN, D., 1998. CONFIDENCE INTERVALS FOR THE NUMBER NEEDED TO TREAT. BMJ, 317(7168), PP. 1309-1312.

ARKEMA, E., VAN VOLLENHOVEN, R. & ASKLING, J., 2012. INCIDENCE OF PROGRESSIVE MULTIFOCAL LEUKOENCEPHALOPATHY IN PATIENTS WITH RHEUMATOID ARTHRITIS: A NATIONAL POPULATION-BASED STUDY. ANN RHEUM DIS, 71(11), PP. 1865-1867.

BLOOMGREN, G. ET AL., 2012. RISK OF NATALIZUMAB-ASSOCIATED PROGRESSIVE MULTIFOCAL LEUKOENCEPHALOPATHY. N ENGL J MED, 366(20), PP. 1870-1880.

CALABRESI, P. ET AL., 2014. SAFETY AND EFFICACY OF FINGOLIMOD IN PATIENTS WITH RELAPSING-REMITTING MULTIPLE SCLEROSIS (FREEDOMS II): A DOUBLE-BLIND, RANDOMISED, PLACEBO-CONTROLLED, PHASE 3 TRIAL. LANCET NEUROL, 13(6), PP. 545-556.

CAUCHETEUX, N. ET AL., 2015. CRITERIA IMPROVING MULTIPLE SCLEROSIS DIAGNOSIS AT THE FIRST MRI. J NEUROL, 262(4), PP. 979-987.

CITROME, L., 2010. RELATIVE VS. ABSOLUTE MEASURES OF BENEFIT AND RISK: WHAT'S THE DIFFERENCE? ACTA PSYCHIATR SCAND, 121(2), PP. 94-102.

CITROME, L. & KETTER, T., 2013. WHEN DOES A DIFFERENCE MAKE A DIFFERENCE? INTERPRETATION OF NUMBER NEEDED TO TREAT, NUMBER NEEDED TO HARM, AND LIKELIHOOD TO BE HELPED OR HARMED. INT J CLIN PRACT, 67(5), PP. 407-411.

COHEN, J. ET AL., 2012. ALEMTUZUMAB VERSUS INTERFERON BETA 1A AS FIRST-LINE TREATMENT FOR PATIENTS WITH RELAPSING-REMITTING MULTIPLE SCLEROSIS: A RANDOMISED CONTROLLED PHASE 3 TRIAL. LANCET, 380(9856), PP. 1819-1828.

COLES, A. ET AL., 2012. ALEMTUZUMAB FOR PATIENTS WITH RELAPSING MULTIPLE SCLEROSIS AFTER DISEASE-MODIFYING THERAPY: A RANDOMISED CONTROLLED PHASE 3 TRIAL. LANCET, 380(9856), PP. 1829-1839.

COOK, R. & SACKETT, D., 1995. THE NUMBER NEEDED TO TREAT: A CLINICALLY USEFUL MEASURE OF TREATMENT EFFECT. BMJ, 310(6977), PP. 452-454.

DERSIMONIAN, R. & LAIRD, N., 1986. META-ANALYSIS IN CLINICAL TRIALS. CONTROL CLIN TRIALS. 1986, 7(3), PP. 177-188.

EMA, 2007. COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE (CHMP). REFLECTION PAPER ON BENEFIT-RISK ASSESSMENT METHODS IN THE CONTEXT OF THE EVALUATION OF MARKETING AUTHORISATION APPLICATIONS OF MEDICINAL PRODUCTS FOR HUMAN USE (EMEA/CHMP/15404/2007). [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/REGULATORY_AND_PROCEDURAL_GUIDELINE/2010/01/WC500069634.PDF [ACCESSED 10 MAY 2016].

EMA, 2015C. NEW RECOMMENDATIONS TO MINIMISE RISKS OF THE RARE BRAIN INFECTION PML AND A TYPE OF SKIN CANCER WITH GILENYA [PRESS RELEASE: 18/12/2015]. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/PRESS_RELEASE/2015/12/WC500199041.PDF [ACCESSED 13 MAY 2016].

EMA, 2015D. UPDATED RECOMMENDATIONS TO MINIMISE THE RISK OF THE RARE BRAIN INFECTION PML WITH TECFIDERA [PRESS RELEASE: 23/10/2015]. [ONLINE] AVAILABLE AT: HTTP://WWW.EMA.EUROPA.EU/DOCS/EN_GB/DOCUMENT_LIBRARY/PRESS_RELEASE/2015/10/WC500196017.PDF [ACCESSED 13 MAY 2016].

ENGLISH, C. & ALOI, J., 2015. NEW FDA-APPROVED DISEASE-MODIFYING THERAPIES FOR MULTIPLE SCLEROSIS. CLIN THER, 37(4), PP. 691-715.

FAULKNER, M., 2015. RISK OF PROGRESSIVE MULTIFOCAL LEUKOENCEPHALOPATHY IN PATIENTS WITH MULTIPLE SCLEROSIS.. EXPERT OPIN DRUG SAF, 14(11), PP. 1737-1748.

FDA, 2013A. STRUCTURED APPROACH TO BENEFIT-RISK ASSESSMENT IN DRUG REGULATORY DECISION-MAKING: DRAFT PATIENT DRUG USER FREE ACT (PDUFA) I IMPLEMENTATION PLAN. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DOWNLOADS/FORINDUSTRY/USERFEES/PRESCRIPTIONDRUGUSERFEE/UCM329758.PDF [ACCESSED 02 APRIL 2015].

FOX, R. ET AL., 2012. PLACEBO-CONTROLLED PHASE 3 STUDY OF ORAL BG-12 OR GLATIRAMER IN MULTIPLE SCLEROSIS. N ENGL J MED, 367(12), PP. 1087-1097.

FRANCIS, G., 2004. IMPORTANCE OF BENEFIT-TO-RISK ASSESSMENT FOR DISEASE-MODIFYING DRUGS USED TO TREAT MS. J NEUROL, 251(SUPPL 5), PP. V42-V49.

FURUKAWA, T., GUYATT, G. & GRIFFITH, L., 2002. CAN WE INDIVIDUALIZE THE 'NUMBER NEEDED TO TREAT'? AN EMPIRICAL STUDY OF SUMMARY EFFECT MEASURES IN META-ANALYSES. INT J EPIDEMIOL, 31(1), PP. 72-76.

GILENYA, 2016A. SUMMARY OF PRODUCT CHARACTERISTICS. DATE OF FIRST AUTHORISATION: 17 MARCH 2011. DATE OF LATEST RENEWAL: 23 NOVEMBER 2015. REVISED: 07/03/2016, S.L.: NOVARTIS EUROPHARM LIMITED.

GILENYA, 2016B. PRESCRIBING INFORMATION. NOVARTIS AG. INITIAL U.S. APPROVAL: 2010. REVISED: 2/2016, S.L.: NOVARTIS AG.

GILENYAWORLDWATCH.COM, 2016. GILENYA WORLD WATCH. [ONLINE] AVAILABLE AT: HTTP://GILENYAWORLDWATCH.COM/ENGLISH.HTML [ACCESSED 13 MAY 2016].

GOLD, R. ET AL., 2012. PLACEBO-CONTROLLED PHASE 3 STUDY OF ORAL BG-12 FOR RELAPSING MULTIPLE SCLEROSIS. N ENGL J MED, 367(12), PP. 1098-1107.

HADJIGEORGIOU, G. ET AL., 2013. A NETWORK META-ANALYSIS OF RANDOMIZED CONTROLLED TRIALS FOR COMPARING THE EFFECTIVENESS AND SAFETY PROFILE OF TREATMENTS WITH MARKETING AUTHORIZATION FOR RELAPSING MULTIPLE SCLEROSIS. J CLIN PHARM THER, 38(6), PP. 433-439.

HAMMAD, T. ET AL., 2013. THE FUTURE OF POPULATION-BASED POSTMARKET DRUG RISK ASSESSMENT: A REGULATOR'S PERSPECTIVE. CLIN PHARMACOL THER, 94(3), PP. 349-358.

HAMMAD, T. ET AL., 2013-B. REPORTING OF META-ANALYSES OF RANDOMIZED CONTROLLED TRIALS WITH A FOCUS ON DRUG SAFETY: AN EMPIRICAL ASSESSMENT. CLIN TRIALS, 10(3), PP. 389-397.

HIGGINS, J., THOMPSON, S., DEEKS, J. & ALTMAN, D., 2003. MEASURING INCONSISTENCY IN META-ANALYSES. BMJ, 327(7414), PP. 557-560.

HUTCHINSON, M. ET AL., 2014. EFFICACY AND SAFETY OF BG-12 (DIMETHYL FUMARATE) AND OTHER DISEASE-MODIFYING THERAPIES FOR THE TREATMENT OF RELAPSING-REMITTING MULTIPLE SCLEROSIS: A SYSTEMATIC REVIEW AND MIXED TREATMENT COMPARISON. CURR MED RES OPIN, 30(4), PP. 613-627.

INGWERSEN, J., AKTAS, O. & HARTUNG, H., 2016. ADVANCES IN AND ALGORITHMS FOR THE TREATMENT OF RELAPSING-REMITTING MULTIPLE SCLEROSIS. NEUROTHERAPEUTICS, 13(1), PP. 47-57.

JACOBS, L. ET AL., 1996. INTRAMUSCULAR INTERFERON BETA-1A FOR DISEASE PROGRESSION IN RELAPSING MULTIPLE SCLEROSIS. THE MULTIPLE SCLEROSIS COLLABORATIVE RESEARCH GROUP (MSCRG). ANN NEUROL, 39(3), PP. 285-294.

JOHNSON, K. ET AL., 1995. COPOLYMER 1 REDUCES RELAPSE RATE AND IMPROVES DISABILITY IN RELAPSING-REMITTING MULTIPLE SCLEROSIS: RESULTS OF A PHASE III MULTICENTER, DOUBLE-BLIND PLACEBO-CONTROLLED TRIAL. THE COPOLYMER 1 MULTIPLE SCLEROSIS STUDY GROUP. NEUROLOGY, 45(7), PP. 1268-1276.

KAPPOS, L. ET AL., 2010. A PLACEBO-CONTROLLED TRIAL OF ORAL FINGOLIMOD IN RELAPSING MULTIPLE SCLEROSIS. N ENGL J MED, 362(5), PP. 387-401.

KLAWITER, E., CROSS, A. & NAISMITH, R., 2009. THE PRESENT EFFICACY OF MULTIPLE SCLEROSIS THERAPEUTICS: IS THE NEW 66% JUST THE OLD 33%? NEUROLOGY, 73(12), PP. 984-990.

LAUPACIS, A., SACKETT, D. & ROBERTS, R., 1988. AN ASSESSMENT OF CLINICALLY USEFUL MEASURES OF THE CONSEQUENCES OF TREATMENT. N ENGL J MED, 318(26), PP. 1728-1733.

MCGUIGAN, C. ET AL., 2016. STRATIFICATION AND MONITORING OF NATALIZUMAB-ASSOCIATED PROGRESSIVE MULTIFOCAL LEUKOENCEPHALOPATHY RISK: RECOMMENDATIONS FROM AN EXPERT GROUP. J NEUROL NEUROSURG PSYCHIATRY, PP. 117-125.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2015. NUMBER NEEDED TO HARM IN THE POST-MARKETING SAFETY EVALUATION: RESULTS FOR ROSIGLITAZONE AND PIOGLITAZONE. PHARMACOEPIDEMIOL DRUG SAF, 24(12), PP. 1259-1270.

MOHER, D., LIBERATI, A., TETZLAFF, J. & ALTMAN, D., 2009. PREFERRED REPORTING ITEMS FOR SYSTEMATIC REVIEWS AND META-ANALYSES: THE PRISMA STATEMENT. ANN INTERN MED, 151(4), PP. 264-269.

MT-ISA, S. ET AL., 2014. BALANCING BENEFIT AND RISK OF MEDICINES: A SYSTEMATIC REVIEW AND CLASSIFICATION OF AVAILABLE METHODOLOGIES. PHARMACOEPIDEMIOL DRUG SAF, 23(7), PP. 667-678.

NIEUWKAMP, D. ET AL., 2015. PML IN A PATIENT WITHOUT SEVERE LYMPHOCYTOPENIA RECEIVING DIMETHYL FUMARATE. N ENGL J MED, 372(15), PP. 1474-1476.

O'CONNOR, P. ET AL., 2009. 250 MICROG OR 500 MICROG INTERFERON BETA-1B VERSUS 20 MG GLATIRAMER ACETATE IN RELAPSING-REMITTING MULTIPLE SCLEROSIS: A PROSPECTIVE, RANDOMISED, MULTICENTRE STUDY. LANCET NEUROL, 8(10), PP. 889-897.

O'CONNOR, P. ET AL., 2011. TEMSO TRIAL GROUP. RANDOMIZED TRIAL OF ORAL TERIFLUNOMIDE FOR RELAPSING MULTIPLE SCLEROSIS. N ENGL J MED, 365(14), PP. 1293-1303.

OH, J. & O'CONNOR, P., 2014. TERIFLUNOMIDE IN THE TREATMENT OF MULTIPLE SCLEROSIS: CURRENT EVIDENCE AND FUTURE PROSPECTS. THER ADV NEUROL DISORD, 7(5), PP. 239-252.

PATY, D. & LI, D., 1993. INTERFERON BETA-1B IS EFFECTIVE IN RELAPSING-REMITTING MULTIPLE SCLEROSIS. II. MRI ANALYSIS RESULTS OF A MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED TRIAL. UBC MS/MRI STUDY GROUP AND THE IFNB MULTIPLE SCLEROSIS STUDY GROUP. NEUROLOGY, 43(4), PP. 662-667.

PATY, D. & LI, D., 2001. INTERFERON BETA-LB IS EFFECTIVE IN RELAPSING-REMITTING MULTIPLE SCLEROSIS. II. MRI ANALYSIS RESULTS OF A MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED TRIAL. 1993 [CLASSICAL ARTICLE]. NEUROLOGY, 57(12 SUPPL 5), PP. S10-S5.

PLAVINA, T. ET AL., 2014. ANTI-JC VIRUS ANTIBODY LEVELS IN SERUM OR PLASMA FURTHER DEFINE RISK OF NATALIZUMAB-ASSOCIATED PROGRESSIVE MULTIFOCAL LEUKOENCEPHALOPATHY. ANN NEUROL, 76(6), PP. 802-812.

POLMAN, C. ET AL., 2006. A RANDOMIZED, PLACEBO-CONTROLLED TRIAL OF NATALIZUMAB FOR RELAPSING MULTIPLE SCLEROSIS. N ENGL J MED, 354(9), PP. 899-910.

RAWLINS, M., 2008. DE TESTIMONIO: ON THE EVIDENCE FOR DECISIONS ABOUT THE USE OF THERAPEUTIC INTERVENTIONS. LANCET, 372(9656), PP. 2152-2161.

ROSKELL, N. ET AL., 2012. ANNUALIZED RELAPSE RATE OF FIRST-LINE TREATMENTS FOR MULTIPLE SCLEROSIS: A META-ANALYSIS, INCLUDING INDIRECT COMPARISONS VERSUS FINGOLIMOD. CURR MED RES OPIN, 28(5), PP. 767-780.

SHEREMATA, W., BROWN, A. & RAMMOHAN, K., 2015. DIMETHYL FUMARATE FOR TREATING RELAPSING MULTIPLE SCLEROSIS. EXPERT OPIN DRUG SAF, 14(1), PP. 161-170.

STRAUS, S., GLASZIOU, P., RICHARDSON, W. & HAYNES, R., 2011. EVIDENCE-BASED MEDICINE: HOW TO PRACTICE AND TEACH IT. 4TH ED. LONDON, UK: CHURCHILL LIVINGSTONE.

TRAMACERE, I. ET AL., 2015. IMMUNOMODULATORS AND IMMUNOSUPPRESSANTS FOR RELAPSING-REMITTING MULTIPLE SCLEROSIS: A NETWORK META-ANALYSIS. COCHRANE DATABASE SYST REV, VOLUME 9:CD011381.

TUTTLE, A. ET AL., 2015. INCREASING PLACEBO RESPONSES OVER TIME IN U.S. CLINICAL TRIALS OF NEUROPATHIC PAIN. PAIN, 156(12), PP. 2616-2626.

TYSABRI, 2012. PRESCRIBING INFORMATION. BIOGEN IDEC INC. INITIAL U.S. APPROVAL: 2004. REVISED: 2/2012, S.L.: BIOGEN IDEC INC.

TYSABRI, 2016. SUMMARY OF PRODUCT CHARACTERISTICS. BIOGEN IDEC LIMITED. DATE OF FIRST AUTHORISATION: 27TH JUNE 2006. REVISED: 12/05/2016., S.L.: BIOGEN IDEC LIMITED.

VERMERSCH, P. ET AL., 2014. TERIFLUNOMIDE VERSUS SUBCUTANEOUS INTERFERON BETA-1A IN PATIENTS WITH RELAPSING MULTIPLE SCLEROSIS: A RANDOMISED, CONTROLLED PHASE 3 TRIAL. MULT SCLER, 20(6), PP. 705-716.

YUAN, Z., LEVITAN, B. & BERLIN, J., 2011. BENEFIT-RISK ASSESSMENT: TO QUANTIFY OR NOT TO QUANTIFY, THAT IS THE QUESTION. PHARMACOEPIDEMIOL DRUG SAF, 20(6), PP. 653-656.

IV.7. SUPPLEMENTAL DATA IV

IV.7.1. SUPPLEMENTAL DATA IV.1 - SEARCH STRATEGY

Supplemental Table IV. 1 – PubMed and Cochrane Central Register of Controlled Trials (CENTRAL) Search Strategy for Randomized Controlled Trials (Searches Conducted on May 10, 2016).

	Search Terms	PubMed	Cochrane
#1	randomized controlled trial[pt]	410904	397495
#2	controlled clinical trial[pt]	496376	395027
#3	randomized[tiab]	375680	331637
#4	placebo[tiab]	175205	166068
#5	clinical trials as topic[mesh:noexp]	175372	34556
#6	randomly[tiab]	251862	132842
#7	trial[ti]	151280	166161
#8	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7	1027421	698122
#9	animals[mh] NOT humans[mh]	4208738	6219
#10	#8 NOT #9	947214	692065
#11	multiple sclerosis[mh]	47532	2131
#12	#10 AND #11	3978	1850
#13	interferon-beta[mh] OR avonex OR belerofon OR betaferon OR rebif OR betaseron OR extavia	8334	709
#14	glatiramer acetate[mh] OR glatiramer OR copaxone OR co*polymer*1 OR "copolymer 1" OR "co polymer 1" OR "cop 1"	39144	482
#15	natalizumab[mh] OR natalizumab OR tysabri OR antegren OR "anti vla4" OR "anti-vla 4" OR "anti alpha4 integrin" OR "an100226" OR "an 100226"	2014	233
#16	fingolimod hydrochloride[mh] OR fingolimod OR fty720 OR "fty 720" OR gilenya	2102	282
#17	alemtuzumab[nm] OR alemtuzumab OR campath OR mabcampath OR "ldp 103" OR ldp103 OR lemtrada	2653	401
#18	teriflunomide[nm] OR teriflunomide OR "a 771726" OR "a77 1726" OR a771726 OR "hmr 1726" OR hmr1726 OR "rs 61980" OR rs61980 OR "su 0020" OR su0020 or aubagio	460	126
#19	dimethyl fumarate[mh] OR "bg 00012" OR bg00012 OR "bg 12" OR "brn 0774590" OR panaclar OR tecfidera	1887	110
#20	peginterferon beta-1a [nm] OR "pegylated interferon beta" OR "plegridy"	41	17
#21	daclizumab HYP[nm] OR daclizumab [nm] OR daclizumab OR zenapax OR zinbryta	1000	429
#22	#13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21	55555	2408
#23	#12 AND #22	1244	538

IV.7.2. SUPPLEMENTAL DATA IV.2 - INCLUSION AND EXCLUSION CRITERIA OF STUDIES

Supplemental Table IV. 2 – Inclusion and exclusion criteria of studies.

The inclusion and exclusion criteria were as follows: (i) phase III randomized clinical trials (RCTs), blinded (single or double), controlled with placebo or active comparators, and treatment duration ≥2 years (≥96 weeks); (ii) adult patients (aged ≥18 years old) with a confirmed diagnosis of RRMS, according to the McDonald criteria or the revised McDonald criteria; (iii) monotherapy with a currently approved DMT, namely alemtuzumab (12 mg/day intravenously [IV], for 2 treatment courses: the first for 5 consecutive days, and the second [12 months later] for 3 consecutive days), DMF (240 mg oral, twice daily), fingolimod (0.5 mg oral, once daily), GA (20 mg subcutaneous [SC], once daily), interferon[IFN]-β -1a (30 mcg intramuscular [IM], once weekly), IFN-β -1a (44 mcg SC, three times a week), IFN-β -1b (250 mcg SC, once every 2 days), natalizumab (300 mg IV, once every 4 weeks), peginterferon-β -1a (125 mcg SC, once every 2 weeks), and teriflunomide (14 mg oral, once daily); (iv) at least 100 patients randomized in every arm of the study. (iv) studies and treatment arms involving non-licensed doses were excluded; for example, of the two studied doses of fingolimod (1.25 mg daily and 0.5 mg daily), only the fingolimod 0.5 mg arm was included in the analysis as it is the licensed dose in Europe. (v) dose-comparison studies were excluded.

IV.7.3. SUPPLEMENTAL DATA IV.3 – CHARACTERISTICS OF DISEASE-MODIFYING THERAPIES

Supplemental Table IV. 3 – Therapeutic indications, contraindications, special warnings and precautions for use, and most common adverse events with disease-modifying therapies included in the study.

DMT	Therapeutic indications approved in European SPC*	Contraindications	Special warnings and precautions for use	Most common adverse reactions
First-line therapies				
IFN beta-1b (Betaseron®/ Betaferon®/ Extavia®)	Betaferon is indicated for the treatment of patients with RRMS and two or more relapses within the last two years.	Hypersensitivity to natural or recombinant interferon beta, human albumin or to any of the excipients; Current severe depression and/or suicidal ideation; Decompensated liver disease; Treatment initiation in pregnancy.	Anaphylaxis and Other Allergic Reactions; Congestive Heart Failure (CHF); Depression and Suicide; Flu-like Symptom Complex; Hepatic Injury; Injection Site Reactions (ISRs) including Necrosis; Leukopenia; Seizures; Thrombotic Microangiopathy; Monitoring for Laboratory Abnormalities (blood cell counts [BCC], liver function).	Flu-like symptoms (fever, chills, arthralgia, malaise, sweating, headache, or myalgia), and ISRs (redness, swelling, discoloration, inflammation, pain, hypersensitivity, necrosis and non-specific reactions).
IFN beta-1a IM (Avonex®)	AVONEX is indicated for the treatment of patients diagnosed with RRMS. In clinical trials, this was characterized by two or more acute exacerbations (relapses) in the previous three years without evidence of continuous progression between relapses.	Hypersensitivity to natural or recombinant interferon beta, human albumin or to any of the excipients; Current severe depression and/or suicidal ideation; Treatment initiation in pregnancy.	Anaphylaxis and Other Allergic-Reactions; Autoimmune Disorders; CHF; Decreased Peripheral Blood Counts; Depression, Suicide, and Psychotic Disorders; Hepatic Injury; Thrombotic Microangiopathy; Monitoring for Laboratory Abnormalities (BCC, liver function).	Flu-like symptoms (myalgia, fever, chills, sweating, asthenia, headache and nausea), and ISRs
IFN beta-1a SC (Rebif®)	Rebif is indicated in patients with relapsing multiple sclerosis. In clinical trials, this was characterized by two or more acute exacerbations in the previous two years	Hypersensitivity to natural or recombinant interferon beta or to any of the excipients; Current severe depression and/or suicidal ideation; Treatment initiation in pregnancy.	Anaphylaxis and Other Allergic Reactions; Decreased Peripheral Blood Counts; Depression and Suicide; Hepatic Injury; ISRs including Necrosis; Seizures; Thrombotic Microangiopathy; Monitoring for Laboratory Abnormalities (BCC, liver function).	Flu-like symptoms, ISRs (predominantly mild inflammation or erythema), and asymptomatic increases in laboratory parameters of hepatic function and decreases in white blood cells.
Glatiramer acetate (Copaxone®)	COPAXONE is indicated for the treatment of patients with relapsing-forms of multiple sclerosis.	Hypersensitivity to glatiramer acetate or mannitol; Pregnant women.	Chest Pain; Immediate Post-Injection Reaction (IPIR); Convulsions and/or anaphylactoid or allergic reactions; Serious hypersensitivity reactions; Lipoatrophy and Skin Necrosis; Potential Effects on Immune Response (glatiramer acetate-reactive antibodies); Monitoring for Laboratory Abnormalities (renal function).	ISRs (erythema, pain, mass, pruritus, oedema, inflammation and hypersensitivity, and rare occurrences of lipoatrophy and skin necrosis), and IPIR (vasodilatation [flushing], chest pain, dyspnoea, palpitation or tachycardia).
Dimethyl fumarate (Tecfidera®)	Tecfidera is indicated for the treatment of adult patients with RRMS.	Hypersensitivity to the active substance or to any of the excipients.	Anaphylaxis and Angioedema; Flushing; Lymphopenia; Progressive Multifocal Leukoencephalopathy (PML); Monitoring for Laboratory Abnormalities (BCC, liver function, renal function).	Flushing and gastrointestinal events (diarrhoea, nausea, abdominal pain, abdominal pain upper).
Teriflunomide (Aubagio®)	AUBAGIO is indicated for the treatment of adult patients with RRMS.	Hypersensitivity to the active substance or to any of the excipients; Breast-feeding women;	Bone Marrow Effects/ Immunosuppression Potential/ Infections; Concomitant Use with Immunosuppressive or Immunomodulating Therapies (has not been evaluated); Hepatotoxicity;	Headache, diarrhoea, increased ALT, nausea, and alopecia.

DMT	Therapeutic indications approved in European SPC*	Contraindications	Special warnings and precautions for use	Most common adverse reactions
		Pregnant women, or women of childbearing potential; Severe active infection; Severe hepatic impairment (Child-Pugh class C); Severe hypoproteinaemia (e.g. nephrotic syndrome); Severe immunodeficiency states (e.g. AIDS); Severe renal impairment undergoing dialysis; Significantly impaired bone marrow function or significant anaemia, leukopenia, neutropenia or thrombocytopenia.	Increased Blood Pressure; Peripheral Neuropathy; Procedure for Accelerated Elimination of Teriflunomide; Respiratory Effects (interstitial lung diseases); Skin Reactions; Use in Women of Childbearing Potential; Monitoring for Laboratory Abnormalities (blood pressure, ALT, BCC).	
Second-line therapies or highly active RRMS				
Natalizumab (Tysabri®)	TYSABRI is indicated as single disease modifying therapy in highly active RRMS for the following patient groups: - Adult patients aged 18 years and over with high disease activity despite treatment with a beta-interferon or glatiramer acetate. - Adult patients aged 18 years and over with rapidly evolving severe relapsing remitting multiple sclerosis.	Hypersensitivity to natalizumab or to any of the excipients; Active malignancies, except for patients with cutaneous basal cell carcinoma; Children and adolescents below the age of 18 years; Combination with beta-interferons or glatiramer acetate; Increased risk for opportunistic infections, including immunocompromised patients; PML.	Hepatotoxicity; Herpes Encephalitis and Meningitis; Hypersensitivity/ Antibody Formation; Immunosuppression/ Infections; PML; Monitoring for Laboratory Abnormalities (BCC, MRI scans, liver function).	Dizziness, nausea, urticaria, and rigors.
Fingolimod (Gilenya®)	Gilenya is indicated as single disease modifying therapy in highly active RRMS for the following adult patient groups: - Patients with highly active disease despite a full and adequate course of treatment with at least one disease modifying therapy, or - Patients with rapidly evolving severe relapsing remitting multiple sclerosis defined by 2 or more disabling relapses in one year, and with 1 or more Gadolinium enhancing lesions on brain. MRI or a significant increase in T2 lesion load as compared to a previous recent MRI.	Hypersensitivity to the active substance or to any of the excipients; Active malignancies; Immunodeficiency syndrome; Increased risk for opportunistic infections, including immunocompromised patients; Severe active infections, including active chronic infections (hepatitis, tuberculosis); Severe liver impairment (Child-Pugh class C);	Basal Cell Carcinoma; Bradyarrhythmia and Atrioventricular Blocks; Foetal Risk; Hypersensitivity Reaction; Immune System Effects Following Discontinuation; Increased Blood Pressure; Infections; Liver Injury; Macular Oedema; Posterior Reversible Encephalopathy Syndrome; PML; Respiratory Effects; Monitoring for Laboratory Abnormalities (BCC, liver function, bradycardia monitoring for 6 hours after first dose, ophthalmological examination).	Influenza, sinusitis, headache, diarrhoea, back pain, hepatic enzyme increased and cough.
Alemtuzumab (Lemtrada®)	LEMTRADA is indicated for adult patients with RRMS with active disease defined by clinical or	Hypersensitivity to the active substance, or to any of the excipients;	Autoimmune cytopenias; Autoimmunity; Glomerular nephropathies; Immune thrombocytopenia;	Rash, headache, pyrexia, and respiratory tract infections.

DMT	Therapeutic indications approved in European SPC*	Contraindications	Special warnings and precautions for use	Most common adverse reactions
	imaging features.	HIV infection.	Infections; Infusion reactions; Malignancies; Thyroid disorders; Monitoring for laboratory abnormalities (BCC, thyroid function tests, serum creatinine, urinalysis with microscopy)	

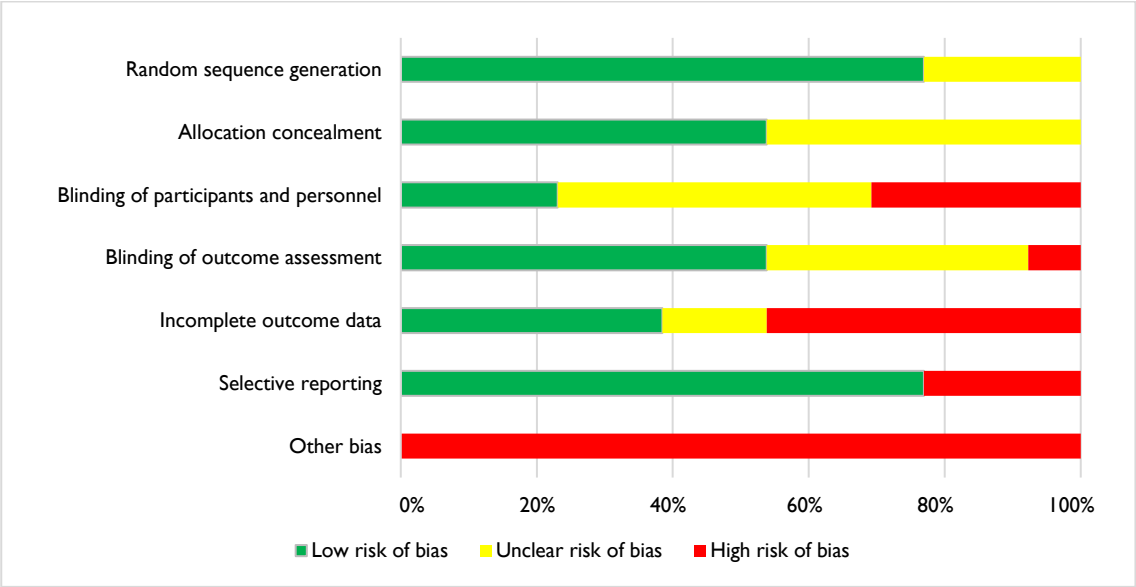
AIDS, acquired immune deficiency syndrome; ALT, alanine aminotransferase; BCC, blood cell counts; CHF, congestive heart failure; DMT, disease-modifying treatment; HIV, human immunodeficiency virus; IAR, infusion-associated reaction; IPIR, immediate post-injection reaction; ISR, injection-site reaction; ITP, immune thrombocytopenic purpura; MRI, magnetic resonance imaging; PML, progressive multifocal leukoencephalopathy; RRMS, relapsing remitting multiple sclerosis; SPC, summary of product characteristics.

IV.7.4. SUPPLEMENTAL DATA IV.4 – ADVERSE EVENTS OF INTEREST

Supplemental Table IV. 4 – List of adverse events of interest selected for each disease-modifying therapy.

Disease-modifying therapy	Adverse events of interest
Dimethyl fumarate	Abdominal pain, diarrhoea, flushing, abnormal liver function (e.g., increase in alanine aminotransferase [ALT]), lymphopenia, and nausea;
Glatiramer acetate	Chest pain, dyspnoea, immediate post-injection reaction (IPIR), and injection-site reaction (ISR);
Beta-interferons (IFNß -1a-SC, IFNß -1a-IM, and IFNß -1b)	Anaphylactoid/allergic reaction, depression (and/or suicidal tendencies), abnormal liver function, influenza-like symptoms, ISR, and decreased blood counts (anaemia and/or lymphopenia);
Teriflunomide	Abnormal liver function, alopecia, diarrhoea, headache, infections, nausea;
Alemtuzumab	Headache, immune thrombocytopenia, infection, infusion-associated reaction (IAR), rash, and thyroid disorder;
Fingolimod	Abnormal liver function, basal-cell carcinoma, bradycardia, increased BP (or hypertension), infection, and macular oedema;
Natalizumab	Abnormal liver function, anaphylactoid/allergic reaction, infection, lymphocytosis, pneumonia, and rigors.

IV.7.5. SUPPLEMENTAL DATA IV.5 – RISK OF BIAS



Supplemental Figure IV. 1 – Risk of bias graph.

CHAPTER V – NUMBER NEEDED TO TREAT (NNT) IN CLINICAL LITERATURE: AN APPRAISAL

V. NUMBER NEEDED TO TREAT (NNT) IN CLINICAL LITERATURE: AN APPRAISAL

V.1. ABSTRACT

The number needed to treat (NNT) is an absolute effect measure that has been used to assess beneficial and harmful effects of medical interventions. There are several methods that can be used to calculate NNTs, which should be applied depending on different study characteristics, such as the design and type of variable used to measure outcomes. Whether the most recommended methods have been applied to calculate NNTs in studies published in medical literature is yet to be determined. The aim of this study is to assess whether the methods used to calculate NNT in studies published in medical journals are in line with basic methodological recommendations. Top-25 high-impact factor journals in the "General and/or Internal Medicine" category were screened to identify studies assessing pharmacological interventions and reporting NNTs. Studies were categorized according to their design, and type of variables. NNTs were assessed for completeness (baseline risk, time-horizon, and confidence intervals [CI]). The methods used for calculating NNTs in selected studies were compared to basic methodological recommendations published in literature. Data was analysed using descriptive statistics. The search returned 138 citations, 51 were selected. Most were meta-analyses (n=23, 45.1%), followed by clinical trials (n=17, 33.3%), cohort (n=9, 17.6%) and case-control studies (n=2, 3.9%). Binary variables were more common (n=41, 80.4%) than time-to-event (n=10, 19.6%) outcomes. Twenty-six studies (51.0%) reported only NNT to benefit (NNTB), 14 (27.5%) reported both NNTB and NNT to harm (NNTH), and 11 (21.6%) reported only NNTH. Baseline risk (n=37, 72.5%), time-horizon (n=38, 74.5%) and CI (n=32, 62.7%) for NNTs were not always reported. Basic methodological recommendations to calculate NNTs were not followed in 15 studies (29.4%). The proportion of studies applying non recommended methods was particularly high in the case of meta-analyses (n=13, 56.5%). A considerable proportion of studies, particularly meta-analyses, applied methods that are not in line with basic methodological recommendations. Despite their usefulness in assisting clinical decisions, NNTs are uninterpretable if incompletely reported, and may be misleading if calculating methods are inadequate to study designs and variables under evaluation. Further research is needed to confirm present findings.

V.2. INTRODUCTION

The concept of 'number needed to treat' (NNT) was introduced in the medical literature by Laupacis et al. in 1988 (Laupacis, Sackett & Roberts 1988). NNT is an absolute effect measure, which is interpreted as the number of patients needed to be treated with one therapy versus another for one patient to encounter an additional outcome of interest within a defined period of time (Laupacis, Sackett & Roberts 1988), (Cook & Sackett 1995). The computation of NNT is founded on the cumulative incidence of the outcome per number of patients followed over a given period of time, being classically calculated by inverting absolute risk (AR) reduction (also called risk difference [RD]) between two treatment options (Laupacis, Sackett & Roberts 1988), (Cook & Sackett 1995).

There are some characteristics that are inherently associated with the concept of NNT. The resulting value is specific to a single comparison between two treatment options within a single study, rather than an isolated absolute measure of clinical effect of a single intervention. Thus, NNT is specific to the results of a given comparison, not to a particular therapy (McAlister 2008). In addition, three other factors, beyond the efficacy or safety of the intervention and the comparator, influence NNT, namely baseline risk (i.e. control event rate [CER]), time frame, and outcomes (McAlister 2008).

The use of NNT has been valuable in daily clinical practice, namely at assisting physicians in selecting therapeutic interventions (Straus et al. 2011), (Citrome & Ketter 2013). Further, this metric has potential of use as a supportive tool in benefit-risk assessments and help regulators making decisions on drug regulation (Mt-Isa et al. 2014), (Mendes, Alves & Batel-Marques 2015), (Mendes, Alves & Batel-Marques 2016a).

The Consolidated Standards of Reporting Trials (CONSORT) statement recommends the use of both relative and absolute measures of effect for randomized controlled trials (RCTs) with binary and time to event outcomes (Altman et al. 2011), (Moher et al. 2010). The British Medical Journal (BMJ) requires that, whenever possible, absolute rather than relative risks and NNTs with 95% confidence intervals (CIs) are reported in RCTs (BMJ 2016). Yet, few authors express their findings in terms of NNT or AR reduction (Nuovo, Melnikow & Chang 2002), (Hildebrandt, Vervölgyi & Bender 2009), (Alonso-Coello et al. 2016). Relative effect measures, such as relative risk (RR), or odds ratio (OR) are more commonly seen in the scientific literature (Citrome 2010), (Alonso-Coello et al. 2016). Despite the unquestionable usefulness of relative effect measures, they do not reflect baseline risks, making it impracticable to discriminate large from small

treatment effects, and leading sometimes to misleading conclusions (Klawiter, Cross & Naismith 2009), (Citrome 2010), (Mendes, Alves & Batel-Marques 2016b).

Although the NNT has been originally conceived to be used in RCTs (Laupacis, Sackett & Roberts 1988), the concept has been used to express treatment differences in comparative studies with other designs, including systematic reviews and meta-analyses, and observational studies (cohort and case-control studies) (McQuay & Moore 1997), (Furukawa, Guyatt & Griffith 2002), (Moore et al. 2002), (Austin & Laupacis 2011), (Bender & Blettner 2002), (Bender et al. 2007). Noteworthy, the notations 'number needed to treat to benefit' (NNTB), and 'number needed to treat to be harmed' (NNTH) were proposed to distinguish between beneficial and harmful outcomes, respectively (Altman 1998). Furthermore, "number needed to be exposed" (NNE), have been proposed to apply the concept of NNT in observational studies, in which the focus is exposure rather than treatment (Bender & Blettner 2002). NNEB and NNEH can be used to describe the number needed to be exposed for one person to benefit or be harmed (Bender & Blettner 2002). In order to simplify, the term NNT is used throughout this paper.

The calculation of NNT should be based upon the use of methods that align with the characteristics of a given study, such as the research design and the type of variable (e.g. binary, time to event, or continuous) used to express the outcome of interest (Furukawa, Guyatt & Griffith 2002), (Bender & Blettner 2002), (Bender 2005), (Altman & Andersen 1999), (Bjerre & LeLorier 2000), (Bender et al. 2013), (Suissa et al. 2012), (Suissa 2015), (Deeks, Higgins & Altman 2011), (da Costa et al. 2012). The use of inadequate methods may lead to erroneous results (Hildebrandt, Vervölgyi & Bender 2009), (Suissa 2009), (Stang, Poole & Bender 2010), (Suissa et al. 2012), (Suissa 2015). A previous research analysing articles published in four major medical journals found that NNTs were miscalculated in 60% of RCTs involving varying follow-up times (Suissa et al. 2012). The authors of another paper concluded that 50% of the RCTs reporting NNTs derived from time to event outcomes applied inadequate calculation methods (Hildebrandt, Vervölgyi & Bender 2009). Moreover, only 34% of RCTs presented the corresponding CIs for point-estimate NNTs (Hildebrandt, Vervölgyi & Bender 2009). The application of inadequate methods within other research designs, such as using pooled RDs in meta-analyses (Cates 2002), (Smeeth, Haines & Ebrahim 1999), or unadjusted incidence rates in observational studies (Bender & Blettner 2002), (Stang, Poole & Bender 2010) have also been pointed out.

The main goal of this study is to assess whether the methods used to calculate NNT in studies published in medical journals are in line with basic methodological recommendations.

V.3. METHODS

V.3.1. STUDIES REPORTING NNT IN MEDICAL JOURNALS

V.3.1.1. Identification and selection of studies

PubMed was searched for papers reporting NNT estimates that were published between 2006 and 2015 in the top 25 high-impact factor journals in the category of "General and/or Internal Medicine", according to the Science Citation Index (Supplemental Table V. 1 from Supplemental Data V.7.1.) (Thomson-Reuters 2016). The search was restricted to these journals because they are more likely to influence clinicians' perceptions on benefits and harms of medicines (Alves, Batel-Marques & Macedo 2012). No further limits were used in the search strategy (Supplemental Table V. 2 from Supplemental Data V.7.2.).

Titles and abstracts of all retrieved citationswere screened by two independent reviewers (DMM and CCA) to identify potentially relevant publications. Full texts were retrieved for relevant citations. Discrepancies were resolved by majority decision (two of three) involving a third investigator (FBM).

Studies were included if they met the following inclusion criteria: (i) have a control group; (ii) assess the effect of a pharmacological intervention on beneficial and/or harmful outcomes; (iii) express at least one resulting effect by means of the NNT. Studies assessing medical interventions other than pharmacological interventions (e.g., surgical techniques, dietary interventions, lifestyle modifications) were not included.

V.3.1.2. Data extraction

V.3.1.2.1. General characteristics of included studies

Data elements extracted to describe general study characteristics included (i) study reference (authors and journal name); (ii) year of publication; (iii) country (determined by the first author's affiliation); (iv) study design; (v) number of included studies (for systematic reviews and meta-analyses); (vi) number of participants; (vii) study duration (i.e., length of participants' follow-up in longitudinal studies); (viii) disease/condition of the studied population; (ix) pharmacological interventions (including comparators); and (x) primary outcome (including its classification as an efficacy and/or safety outcome). Diseases/conditions were classified using the Medical Dictionary for Regulatory Activities (MedDRA), v. 18.0, according to the System Organ Class (SOC) (Brown, Wood & Wood 1999).

V.3.1.2.2. Characteristics of NNTs in included studies

Data was collected from included studies to describe and characterize NNTs, as well as to allow for further assessment of calculating methods, according to a list of pre-defined queries (Supplemental Table V. 3 and Supplemental Table V. 4 from Supplemental data V.3). When the methodology used to calculate NNTs was not described in the methods section of the included studies, information from the results or the discussion sections, namely statements given in the text, were used to identify the calculating methods.

V.3.2. METHODS RECOMMENDED TO CALCULATE NNT

V.3.2.1. Methodological recommendations

A summary of basic and general recommendations was set up based upon the evidence reported in the Cochrane Handbook for Systematic Reviews of Interventions (Deeks, Higgins & Altman 2011), as well as in a thorough review performed by Bender about methods to obtain NNTs for different study designs (Bender 2005), and also in another review that focused observational studies (Austin & Laupacis 2011). In addition, a limited,

non-systematic literature search was performed in PubMed to identify papers later published that could complement this evidence (Supplemental Table V. 5 from Supplemental data V.4).

V.3.2.1.1. Systematic Review and Meta-Analysis

The NNT should be calculated based upon the use of a relative effect because relative effects tend to be more stable across risk groups than absolute differences (Furukawa, Guyatt & Griffith 2002), (Deeks, Higgins & Altman 2011), (Schmid et al. 1998), (Engels et al. 2000). The RR and OR, obtained within fixed or random effects regression models, appear to be reasonably constant across different baseline risks (Furukawa, Guyatt & Griffith 2002). The pooled RR or OR can be used to calculate individualized NNTs for different baseline risks (i.e. π_0 the risk control group), using formulas (1) or (2) (Furukawa, Guyatt & Griffith 2002), (Bender 2005), (Deeks, Higgins & Altman 2011). Further, expressing RR or OR as a variety of NNTs across a range of different baseline risks has been recommended (McQuay & Moore 1997), (Deeks, Higgins & Altman 2011), (Smeeth, Haines & Ebrahim 1999).

(1) $NNT = \frac{1}{(1-RR)\times\pi_0}, for\ RR < 1;\ NNT = \frac{1}{(RR-1)\times\pi_0}, for\ RR > 1$

(2) $NNT = \frac{1}{(1-OR)\times\pi_0} + \frac{OR}{(1-OR)\times(1-\pi_0)}, for\ OR < 1;\ NNT = \frac{1}{(OR-1)\times\pi_0} + \frac{OR}{(OR-1)\times(1-\pi_0)}, for\ OR > 1$

V.3.2.1.2. Randomized controlled trials

In RCTs with a binary outcome and a defined period of time, during which all patients are followed, the NNT is estimated based upon the use of simple proportions of patients with the outcome (i.e. π_0 the risk control group, and π_1 the risk in treatment group), according to formula (3) (Laupacis, Sackett & Roberts 1988), (Cook & Sackett 1995).

(3) $NNT = \frac{1}{\pi_1-\pi_0} = \frac{1}{RD}$

In RCTs with time to event outcomes, the time of follow-up is not equal for all patients. Simple proportions should not be used to estimate NNTs because they do not account for varying follow-up times (Bender 2005), (Suissa et al. 2012). In such studies, the Kaplan-Meier approach can be used to estimate proportions of patients with the outcome of

interest over time (Altman & Andersen 1999). The NNT can then be calculated by inverting the RD between cumulative incidences (i.e. survival probabilities $S_1(t)$ for treatment groups and $S_0(t)$ for control group) at a given point of time (t), as shown in formula (4) (Altman & Andersen 1999).

$$(4)\ NNT = \frac{1}{S_1(t) - S_0(t)}$$

Further, the hazard ratio (HR), estimated by means of the Cox regression model, can be used to estimate the NNT if the assumption of proportional hazards is fulfilled and $S_0(t)$ is available, as described in formula (5) (Altman & Andersen 1999).

$$(5)\ NNT = \frac{1}{(S_0(t))^{HR} - S_0(t)}$$

V.3.2.1.3. Observational studies

Due to the lack of randomization, the estimation of treatment effects in observational studies requires adjustment for confounding factors (Bender & Blettner 2002). Regression-based methods, namely multiple logistic regression, or propensity score methods can be performed to estimate adjusted relative effects (Austin & Laupacis 2011). The NNT should also be adjusted and not based on crude risk differences without adjustment (Bender & Blettner 2002).

V.3.2.1.3.1. Case-control studies

In case-control studies, multiple logistic regression is usually performed to estimate adjusted OR as relative effect measure (Bender & Blettner 2002), (Bender et al. 2007). The NNT can be calculated by combining the adjusted OR with the risk in control or unexposed group (usually called the unexposed event rate [UER]) (Bender & Blettner 2002), (Bjerre & LeLorier 2000). In case-control studies the UER is obtained from an external source (for example, controls in RCTs or unexposed subjects in cohort studies) (Bjerre & LeLorier 2000). Formula (2), where π_0 = UER, should be used to calculate adjusted NNT from adjusted OR. If the relative effect measure is adjusted RR, then formula (1) should be applied.

V.3.2.1.3.2. Cohort studies

In cohort studies using regression-based methods, two general approaches can be used to estimate NNT. The first approach is based upon the use of adjusted OR, estimated by means of multiple logistic regression (Bender & Blettner 2002). Adjusted NNT is obtained with the application of adjusted OR to UER, as described in formula (2). However, this approach should only be used if there is a small variation of the risks around the mean (Bender et al. 2007). The mean risk of unexposed subjects (UER), which is estimated by means of the logistic regression model, can be used to calculate adjusted NNT for the corresponding confounder profile. Another method that can be used is to calculate NNT for some fixed confounder profiles (Bender & Blettner 2002). In the second approach, NNT is calculated by taking the reciprocal of the average RD over the observed confounder values, estimated by means of multiple logistic regression. In general, the approach based upon the average RD should be applied (Bender et al. 2007).

In case of time to event outcomes, NNT can be estimated as the reciprocal of the difference between two marginal probabilities, within a given duration of follow-up, using an adjusted survival model (e.g. the Cox proportional hazards regression model) (Austin 2010), (Austin & Laupacis 2011), (Laubender & Bender 2010), (Laubender & Bender 2014).

In cohort studies using propensity score methods, NNT can be estimated by inverting RD, which is directly estimated by comparing the probability of the outcome between treated and untreated subjects in the matched sample in propensity-score matching (Austin & Laupacis 2011). If the outcome is time to event, NNT is given by the reciprocal of the difference estimated from Kaplan-Meier survival curves in treated and untreated subjects within a given duration of follow-up (Austin & Laupacis 2011).

V.3.3. ADHERENCE TO METHODOLOGICAL RECOMMENDATIONS

The methods used to calculate NNTs in studies from medical journals were compared to basic methodological recommendations. The adherence of calculating methods to methodological recommendations was assessed, considering the study design, and the type of variable used to measure outcomes of interest.

V.3.4. DATA ANALYSIS

Data were analysed using descriptive statistics. Data analyses were performed using Microsoft® Excel® 2013.

V.4. RESULTS

Figure V. 1 presents the search strategy flowchart. From 138 publications, 51 were selected after excluding studies not fulfilling the inclusion criteria. Table V. 1 presents a summary of the main characteristics of included studies, namely the characteristics of variables and effect measures used to assess effects of interventions, and the completeness of data around NNT estimates. A detailed description of the characteristics of each study is provided in Supplemental data V.5 (Supplemental Table V. 6).

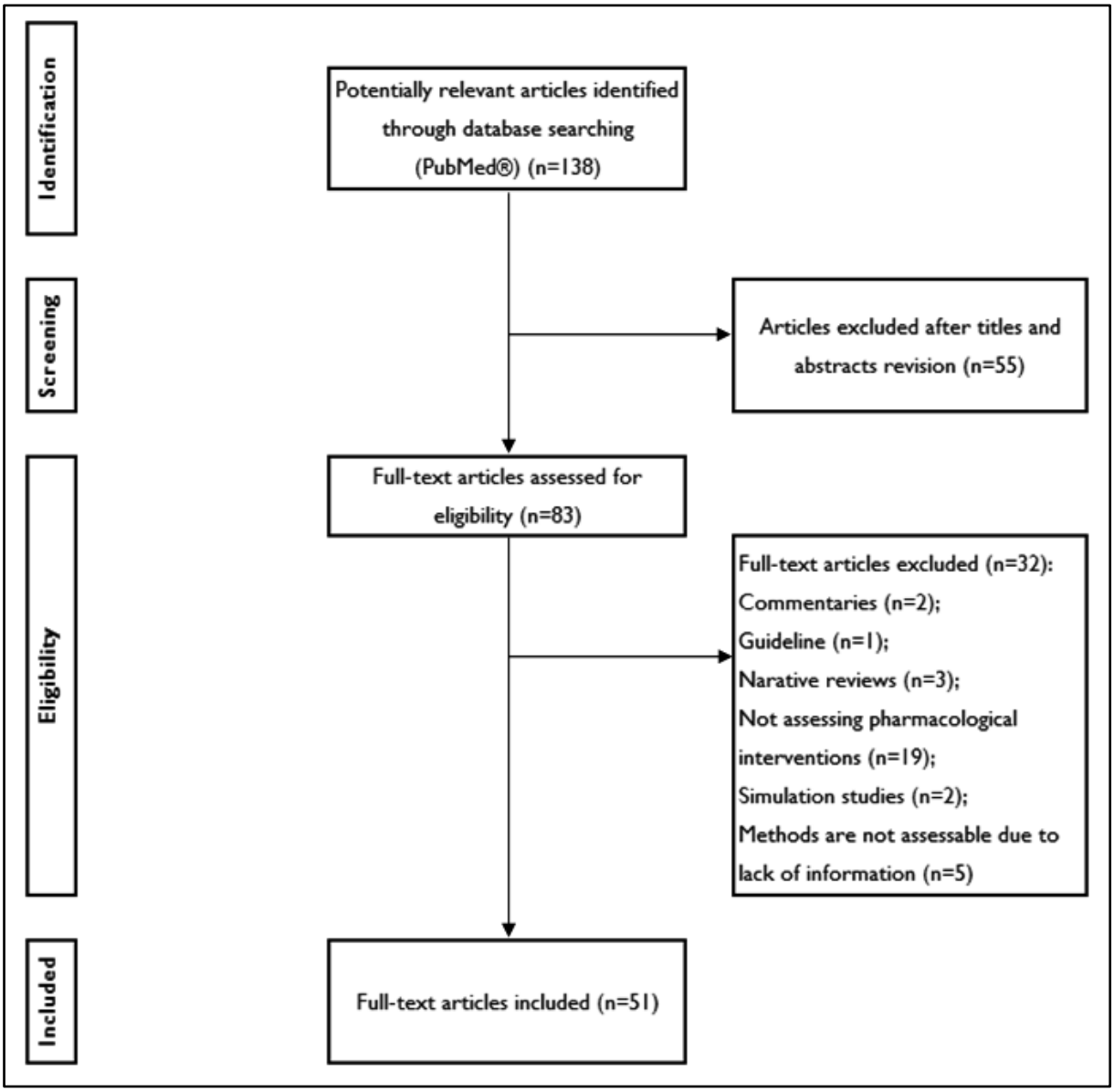


Figure V. 1 – Flow of studies through the review process.

Table V. 1 – Characteristics of the included studies and of the number needed to treat (NNT).

Characteristics	Meta-analysis (n=23)	RCT (n=17)	Cohort (n=9)	Nested case-control (n=2)	Overall (n=51)
Journal					
JAMA	9 (39.1%)	4 (23.5%)	2 (22.2%)	2 (100.0%)	17 (33.3%)
Lancet	6 (26.1%)	7 (41.2%)	1 (11.1%)	0 (0.0%)	14 (27.5%)
Am J Med	2 (8.7%)	0 (0.0%)	2 (22.2%)	0 (0.0%)	4 (7.8%)
Other	6 (26.1%)	6 (35.3%)	4 (44.4%)	0 (0.0%)	16 (31.4%)
Country					
USA	13 (56.5%)	2 (11.8%)	6 (66.7%)	0 (0.0%)	21 (41.2%)
UK	4 (17.4%)	2 (11.8%)	0 (0.0%)	0 (0.0%)	6 (11.8%)
Canada	1 (4.3%)	2 (11.8%)	1 (11.1%)	2 (100.0%)	6 (11.8%)
Other	5 (21.7%)	11 (64.7%)	2 (22.2%)	0 (0.0%)	18 (35.3%)
Disease / condition					
Infections and infestations	4 (17.4%)	2 (11.8%)	1 (11.1%)	0 (0.0%)	7 (13.7%)
Cardiac disorders	3 (13.0%)	3 (17.6%)	1 (11.1%)	0 (0.0%)	7 (13.7%)
Psychiatric disorders	4 (17.4%)	3 (17.6%)	0 (0.0%)	0 (0.0%)	7 (13.7%)
Other	12 (52.2%)	9 (52.9%)	7 (77.8%)	2 (100.0%)	30 (58.8%)
Primary outcome of study					
Efficacy	12 (52.2%)	16 (94.1%)	2 (22.2%)	0 (0.0%)	30 (58.8%)
Safety	2 (8.7%)	1 (5.9%)	6 (66.7%)	2 (100.0%)	11 (21.6%)
Efficacy and Safety	9 (39.1%)	0 (0.0%)	1 (11.1%)	0 (0.0%)	10 (19.6%)
Type of variable (primary outcome)					
Binary	22 (95.7%)†	13 (76.5%)	5 (55.6%)	1 (50.0%)	41 (80.4%)
Time to event	1 (4.3%)	4 (23.5%)	4 (44.4%)	1 (50.0%)	10 (19.6%)
Relative effect measure					
Yes					
Relative Risk	11 (47.8%)‡	5 (29.4%)	2 (22.2%)	0 (0.0%)	18 (35.3%) †
Odds Ratio	9 (39.1%)‡	4 (23.5%)	2 (22.2%)	1 (50.0%)	16 (31.4%) †
Hazard Ratio	1 (4.3%)	3 (17.6%)	3 (33.3%)	0 (0.0%)	7 (13.7%)
Rate Ratio	0 (0.0%)	0 (0.0%)	1 (11.1%)	1 (50.0%)	2 (3.9%)
No	3 (13.0%)	5 (29.4%)	1 (11.1%)	0 (0.0%)	9 (17.6%)
Outcome expressed with NNT					
Primary outcome	6 (26.1%)	14 (82.4%)	7 (77.8%)	1 (50.0%)	28 (54.9%)
Secondary outcome	0 (0.0%)	2 (11.8%)	0 (0.0%)	0 (0.0%)	2 (3.9%)
Primary and secondary outcomes	17 (73.9%)	1 (5.9%)	2 (22.2%)	1 (50.0%)	21 (41.2%)
NNT for benefit or harm?					
Benefit	8 (34.8%)	15 (88.2%)	3 (33.3%)	0 (0.0%)	26 (51.0%)
Harm	2 (8.7%)	1 (5.9%)	6 (66.7%)	2 (100.0%)	11 (21.6%)
Benefit and Harm	13 (56.5%)	1 (5.9%)	0 (0.0%)	0 (0.0%)	14 (27.5%)
Type of NNT calculated in the study					

Characteristics	Meta-analysis (n=23)	RCT (n=17)	Cohort (n=9)	Nested case-control (n=2)	Overall (n=51)
Person-based NNT	21 (91.3%)†	13 (76.5%)	5 (55.6%)	1 (50.0%)	40 (78.4%)
Person-time-based NNT	2 (8.7%)	4 (23.5%)	4 (44.4%)	1 (50.0%)	11 (21.6%)
Completeness of NNT estimate					
Control event rate					
Yes	13 (56.5%)	17 (100.0%)	6 (66.7%)	1 (50.0%)	37 (72.5%)
No	10 (43.5%)	0 (0.0%)	3 (33.3%)	1 (50.0%)	14 (27.5%)
Time horizon					
Yes	10 (43.5%)	17 (100.0%)	9 (100.0%)	2 (100.0%)	37 (72.5%)
No	13 (56.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	14 (27.5%)
Confidence intervals					
Yes	15 (65.2%)§	8 (47.1%)	8 (88.9%)	1 (50.0%)	32 (62.7%)
No	8 (34.8%)	9 (52.9%)	1 (11.1%)	1 (50.0%)	19 (37.3%)

† The variable for the primary outcome of one meta-analysis is binary and pooled OR (95% CI) was calculated. However, a person-time based NNT was calculated by taking the reciprocal of RD between pooled event rates per 1000 patient-years (Preiss et al. 2011).
‡ One single study reported relative risk (RR) and odds ratio (OR) (Maher et al. 2011).
§ Confidence interval was provided with NNT only for the primary outcome in a study reporting NNT for several outcomes (Green et al. 2007).

V.4.1. GENERAL CHARACTERISTICS OF INCLUDED STUDIES

The majority of studies reporting NNTs were identified from the JAMA (n=17; 33.3%), and the Lancet (n=14; 27.5%) (Supplemental Table V. 7 from Supplemental data V.6). The median number of papers per year was 5.5 (ranging from 1 in 2009 to 7 in 2011, 2012 and 2014). The included studies were more frequently authored by researchers from the USA (n=21; 41.2%), UK (n=6; 11.8%), and Canada (n=6; 11.8%).

Twenty-three (45.1%) publications were systematic reviews and meta-analyses, while 17 were individual RCTs (33.3%), 9 cohort studies (17.6%), and 2 case-control studies (3.9%). The more frequently studied diseases/conditions were "infections and infestations" (n=7; 13.7%), "cardiac disorders" (n=7; 13.7%), and "psychiatric disorders (n=7; 13.7%).

The primary outcomes of most studies assessed only efficacy (n=30; 58.8%) of interventions. Safety was assessed as the sole primary outcome in 11 studies (21.6%). The remaining 10 studies (19.6%) assessed both efficacy and safety as a primary outcome. The primary outcome was binary in 41 studies (80.4%) and time to event in 10 studies (19.6%).

In addition to NNT estimates, the majority of studies (n=42; 82.4%) also used relative effect measures to express treatment differences. The RR (n=18; 35.3%) and OR (n=16; 31.4%) were the most commonly used.

V.4.2. CHARACTERISTICS OF NNTS IN INCLUDED STUDIES

NNTs were estimated only for primary outcomes in 28 studies (54.9%), for primary and also secondary outcomes in 21 studies (41.2%), and only for secondary outcomes in 2 studies (3.9%). NNTs were used to assess only benefits of interventions in 26 studies (51.0%), both benefits and harms in 14 studies (27.5%), and only harms in 11 studies (21.6%).

The type of NNT presented in most studies was a person-based NNT (n=40; 78.4%). A person-time-based NNT was presented in 11 studies (21.6%).

The completeness of data presented around the point-estimate NNT was assessed. The baseline risk (i.e. CER) was presented in 37 studies (72.5%), a defined time horizon in 38 studies (74.5%) and CIs in 32 studies (62.7%).

V.4.3. ASSESSMENT OF METHODS USED TO CALCULATE NNTS

Methods used to calculate NNTs in included studies were compared to basic methodological recommendations (Table V. 2). A detailed description of data used to assess the completeness of information and the appropriateness of methods used to compute NNTs in included studies is available in Supplemental data V.6 (Supplemental Table V. 8).

The methodology used to calculate NNT was clearly defined in the methods section of the publications in 28 studies (54.9%). The methodology were not presented in the methods section of the remaining 23 studies (45.1%), but it could be identified using information from other sections of the publications.

Overall, basic methodological recommendations were followed to calculate NNT in 36 studies (70.6%). A summary of the characteristics of studies that did not follow basic methodological recommendations (n=15; 29.4%) is provided in Table V. 3.

NNT was calculated as the inverse of the RD between groups in 39 studies (76.5%) (13 meta-analysis, 17 RCTs and 9 cohort studies). Of those studies, 17 used simple proportions, 12 used pooled RDs, 4 used average RDs, and 6 used cumulative incidence rates. Simple proportions were correctly used in 14 studies (13 RCTs, and 1 cohort study), and inappropriately used in 3 studies (1 meta-analysis, 1 RCT, and 1 cohort study). Pooled RDs were always inadequate to the study design (12 meta-analyses). The average RD method was considered to have been correctly used in all 4 studies (4 cohort studies). Cumulative incidence rates were adequately used in all 6 studies (3 cohort studies, and 3 RCTs).

The result of a relative effect measure (e.g. OR, RR) was applied to a CER to calculate NNT in 12 studies (23.5%) (10 meta-analyses, and 2 case-control studies). The use of this methodology in those studies was in line with basic methodological recommendations.

Table V. 2 – Assessment of methodology used to calculate number needed to treat (NNT) in included studies.

	Meta-analysis (n=23)	RCT (n=17)	Cohort (n=9)	Nested case-control (n=2)	Overall (n=51)
Methodology used to calculate NNT is defined in the methods section of the study					
Yes	19 (82.6%)	0 (0.0%)	7 (77.8%)	2 (100.0%)	28 (54.9%)
No	4 (17.4%)	17 (100.0%)	2 (22.2%)	0 (0.0%)	23 (45.1%)
General characteristics of the methodology used to calculate NNT in the study					
Reciprocal of risk difference					
Simple proportions	1 (4.3%)	14 (82.4%)	2 (22.2%)	0 (0.0%)	17 (33.3%)
Cumulative IR	0 (0.0%)	3 (17.6%)	3 (33.3%)	0 (0.0%)	6 (11.8%)
Pooled RD	12 (52.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	12 (23.5%)
Average RD	0 (0.0%)	0 (0.0%)	4 (44.4%)	0 (0.0%)	4 (7.8%)
Relative effect measure	10 (43.5%)	0 (0.0%)	0 (0.0%)	2 (100.0%)	12 (23.1%)
Methodology used to calculate NNT is in line with basic recommendations (overall)					
Yes	10 (43.5%)	16 (94.1%)	8 (88.9%)	2 (100.0%)	36 (70.6%)
No	13 (56.5%)	1 (5.9%)	1 (11.1%)	0 (0.0%)	15 (29.4%)
Methodology used to calculate NNT is in line with basic recommendations (detailed)					
Binary variables					
Yes	9 (39.1%)	13 (76.5%)	5 (55.6%)	1 (50.0%)	28 (54.9%)
Reciprocal of risk difference					
Simple proportions	0 (0.0%)	13 (76.5%)	1 (11.1%)	0 (0.0%)	14 (27.5%)
Cumulative IR	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Pooled RD	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Average RD	0 (0.0%)	0 (0.0%)	4 (44.4%)	0 (0.0%)	4 (7.8%)
Relative effect measure	9 (39.1%)	0 (0.0%)	0 (0.0%)	1 (50.0%)	10 (19.6%)
No	13 (56.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	13 (25.5%)
Reciprocal of risk difference					
Simple proportions	1 (4.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (2.0%)
Cumulative IR	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Pooled RD	12 (52.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	12 (23.5%)
Average RD	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Relative effect measure	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Time to event variables					
Yes	1 (4.3%)	3 (17.6%)	3 (33.3%)	1 (50.0%)	8 (15.7%)
Reciprocal of risk difference					
Simple proportions	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

	Meta-analysis (n=23)	RCT (n=17)	Cohort (n=9)	Nested case-control (n=2)	Overall (n=51)
Cumulative IR	0 (0.0%)	3 (17.6%)	3 (33.3%)	0 (0.0%)	6 (11.8%)
Pooled RD	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Average RD	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Relative effect measure	1 (4.3%)	0 (0.0%)	0 (0.0%)	1 (50.0%)	2 (3.9%)
No	0 (0.0%)	1 (5.9%)	1 (11.1%)	0 (0.0%)	2 (3.9%)
Reciprocal of risk difference					
Simple proportions	0 (0.0%)	1 (5.9%)	1 (11.1%)	0 (0.0%)	2 (3.9%)
Cumulative IR	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Pooled RD	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Average RD	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Relative effect measure	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

IR, incidence rate; RCT, randomized controlled trial; RD, risk difference.

Table V. 3 – Characteristics of the included studies in which basic recommendations were not followed to calculate the number needed to treat (NNT).

Study	Variable	Baseline risk	Time horizon	Confidence interval	Methodology used to compute NNT defined in methods section	Method used to compute NNT	Source of data used to compute NNT	Comments
Systematic review and meta-analysis								
Jonas 2014	Binary	No	No	Yes	Yes	NNT=1/RD	Pooled RD	A pooled RD was calculated for two outcomes. Duration of included trials ranged from 12 to 52 weeks for the outcome any drinking, and from 12 to 24 weeks for heaving drinking.
Hempel 2012	Binary	No	No	Yes	Yes	NNT=1/RD	Pooled RD	The pooled RD (obtained from meta-analysis) lead to a loss of follow-up time. Most trials either did not specify the follow-up period, or the assessment was explicitly limited to the time of antibiotics treatment.
Leucht 2012	Binary	Yes	Yes	Yes	Yes	NNT=1/RD	Pooled RD	The outcome is assessed between 7 and 12 months of follow-up; a mean study duration is indicated for each outcome with NNT calculated from absolute RD pooled from the meta-analysis.
Shah 2012	Binary	No	No	Yes	Yes	NNT=1/RD	Pooled RD	The study comprehends the calculation and comparison of NNT for several treatments. However, NNTs are not comparable because they were calculated from pooled RDs and times of follow-up vary considerably across studies included in the meta-analysis (10 days to 48 weeks).
Preiss 2011	Binary	Yes	Yes	No	No	NNT=1/RD	Pooled RD	The variable for the primary outcome of the study is binary and pooled OR (95% CI) was calculated. However, NNT was calculated by taking the reciprocal of RD between pooled event rates per 1000 patient-years. Person-time based NNT was presented and interpreted as the number of persons needed to treat over one year.
Shamliyan 2011	Binary	Yes	No	Yes	Yes	NNT=1/RD	Pooled RD	Several antiviral treatments were compared based on estimates NNT. However, studies with different times of follow-up for antiviral treatments were used to pool absolute RD. The time horizon factor is lost.
Coker 2010	Binary	Yes	Yes	Yes	No	NNT=1/RD	Pooled RD	The pooled RD was obtained for a 14 day follow-up duration in every studies included in the meta-analysis. However, RD varies considerably across the studies included in the meta-analysis (ranging from -8% to 27%).
Testa 2008	Binary	No	No	Yes	Yes	NNT=1/RD	Pooled RD	Pooled RD was used to calculate NNT. The follow-up of included studies ranged from 'in hospital' to 6 months.
Bridge 2007	Binary	Yes	No	Yes	Yes	NNT=1/RD	Pooled RD	DerSimonian and Laird random-effects model was used to obtain a pooled estimate of the RD (95% CI). NNT was calculated as the reciprocal of RD. The duration of follow-up and the baseline risk varied considerably across included studies.
Dentali 2007	Binary	Yes	No	No	Yes	NNT=1/RD	Simple proportions	Raw totals of patients from each study were added together to estimate proportions and calculate RD, i.e. treating data as it all came from one study (Simpson's paradox). Further, the baseline risk ranged considerably across included studies (e.g. 0.2% to 4.0% for pulmonary embolism).
Rovers 2006	Binary	Yes	Yes	No	No	NNT=1/RD	Pooled RD	Although it is not clearly stated in the methods section, the discussion of the study suggest that the authors calculated pooled RD by means of the meta-analysis.
Bongartz 2006	Binary	No	Yes	Yes	Yes	NNT=1/RD	Pooled RD	NNT calculated for treatment periods of 6 to 12 months and 3 to 12 months, using Mantel-Haenszel fixed-estimate of absolute RD in cases in which an OR of at least 1.5 was detected.

Study	Variable	Baseline risk	Time horizon	Confidence interval	Methodology used to compute NNT defined in methods section	Method used to compute NNT	Source of data used to compute NNT	Comments
Systematic review and meta-analysis								
Spiegel 2006	Binary	No	No	No	Yes	NNT=1/RD	Pooled RD	A pooled RD was calculated for two comparisons. Duration of included trials ranged from 6 to 78 weeks for one comparison; and from 12 to 24 weeks for another comparison.
Randomized Controlled Trial								
Shepherd 2008	Time to event	Yes	Yes	No	No	NNT=1/RD	Simple proportions	NNT calculated as 1/RD using final rates of event and citing a median time of follow-up of 4.8 years (NNT=14 in patients with diabetes and Chronic Kidney Disease). However, a Kaplan-Meier curve is provided in the study, which should have been used (since the median follow-up is lower than the 5-years objective, at least some patients did not complete the follow-up). From the Kaplan-Meier curve, we would have 20.3% and 14.0% patients with the outcome in the atorvastatin 10 mg and 80 mg/day, respectively, at 4.8 years of follow-up and a NNT = 15.8).
Retrospective Cohort Study								
Graham 2010	Time to event	Yes	Yes	Yes	Yes	NNT=1/RD	Simple proportions	NNT was calculated using RD between unadjusted incidence rates. Adjusted incidence rates from the Kaplan-Meier curves should have been used. For example, at one year of follow-up, NNT for the composite endpoint would be 92 from Kaplan-Meier curves, rather than 60 person-years from unadjusted incidence rates. The authors interpreted person-years as number of persons treated over one year, which is not exactly the same.

V.5. DISCUSSION

The present study provides an overview about the use of the NNT in medical research during last decade. The adherence of selected studies to basic methodological recommendations was reviewed. This topic is particularly relevant given that the NNT concept has been extended to derive related metrics with potential of use in benefit-risk assessments, namely for clinical decision making or drug regulatory purposes. An example is provided by impact numbers, which give a population perspective to the NNT (Heller et al. 2002), (Attia et al. 2002). Impact numbers are useful to describe public health burden of a disease, and the potential impact of a treatment (Mt-Isa et al. 2014). Two measures of impact numbers are particularly interesting, namely the number of events prevented in the population (NEPP) and the population impact number of eliminating a risk factor over time t (PIN-ER-t) (Mt-Isa et al. 2014), (Heller, Edwards & McElduff 2003), (Heller et al. 2003).

Clinicians and other investigators should be aware that the calculation and interpretation of NNTs depend on specific study characteristics, particularly the design and outcome variables. The use of inadequate calculating methods may lead to biased results and misleading conclusions (Bender & Blettner 2002), (Suissa et al. 2012), (Cates 2002), (Hutton 2000).

The majority of studies included in the present review were aimed to assess primarily only efficacy of medical interventions. The NNT was used more often to assess only benefits (51.0%), rather than only harms (21.6%). This finding was expected taking into account what it is commonly seen in the medical literature. A previous systematic review including meta-analyses published over a 5-year period found that only 14% of studies were designed to investigate drug safety as primary outcome (Alves, Batel-Marques & Macedo 2012). In other study comprising systematic reviews with absolute effect estimates, it was found that the NNT was mostly used to assess beneficial outcomes rather than harmful events (Alonso-Coello et al. 2016).

Overall, included studies reported more frequently results for binary outcomes than for time to event outcomes. This finding contrasts with the results of a previous review in which nearly 55% of included studies reported NNTs for time to event outcomes (Hildebrandt, Vervölgyi & Bender 2009). However, that review included only RCTs, while the present study included several research designs.

Relative measures of effect were used to express treatment differences in the majority of included studies (82.4%). These findings are in line with the conclusions of a

recent survey of 202 systematic reviews (Alonso-Coello et al. 2016). Of those, the majority included meta-analyses with estimation of relative effects (92.1%), while absolute effect estimates were provided in 36.1% (Alonso-Coello et al. 2016).

As previously mentioned, the concept of NNT requires the description of a defined period of time, and varies with baseline risk (also called CER). Nevertheless, the time horizon was lacking in more than one fourth (25.5%) of studies. The NNT is uninterpretable if the time of follow-up during which cumulative outcome incidences are measured is not provided (Stang, Poole & Bender 2010). In addition, baseline risks could not be ascertained in nearly 28% of studies. Previous findings indicate that 56.2% of studies reporting absolute risks do not present the source of baseline risk estimates (Alonso-Coello et al. 2016). Lastly, more than one third (37.3%) of studies included in the present review did not report the CI for the point-estimate NNT. This result is in line with previous findings (Hildebrandt, Vervölgyi & Bender 2009). Thus, a moderately high proportion of papers published in journals with high impact factor in the category of "General and/or Internal Medicine" misuse the NNT metric.

As seen across the articles reviewed in here, several approaches have been used to derive NNTs from meta-analyses. However, in 13 out of 23 meta-analyses (56.5%) the approach was considered inadequate, considering basic methodological recommendations. Of these meta-analyses, one calculated the reciprocal of simple proportions (using total numbers of both patients with outcome and exposed patients coming from all included studies). Simple proportions, i.e. treating data as if it all come from a single trial, to calculate NNTs should not be used, as it is prone to bias due to Simpson's paradox (Cates 2002), (Altman & Deeks 2002). The others 12 inverted pooled RDs, but such method should also be avoided (Furukawa, Guyatt & Griffith 2002), (Deeks, Higgins & Altman 2011), (Smeeth, Haines & Ebrahim 1999), (Marx & Bucher 2003). Absolute RDs are usually not constant and homogeneous across different baseline event rates, therefore being rarely appropriate for calculating NNTs from meta-analyses (Furukawa, Guyatt & Griffith 2002), (Deeks, Higgins & Altman 2011), (Smeeth, Haines & Ebrahim 1999), (Marx & Bucher 2003). Moreover, effect of secular trends on disease risk, and time horizon preclude the use of pooled RDs, as they can result in misleading NNTs (Smeeth, Haines & Ebrahim 1999), (Marx & Bucher 2003). Relative effect measures (such as RR and OR) are usually more stable across risk groups than do absolute differences. Thus, pooled estimates of relative effect measures should be used rather than absolute RDs to derive NNTs from meta-analyses (Furukawa, Guyatt & Griffith 2002), (Deeks, Higgins & Altman 2011), (Smeeth, Haines & Ebrahim 1999). Clinicians

should use, preferably, fixed effects OR, random effects OR or RR and the patient expected event rate (PEER) to individualize NNT when applying results from meta-analyses in clinical practice (Furukawa, Guyatt & Griffith 2002), (Straus et al. 2011).

Most RCTs (94.1%) followed basic methodological recommendations to calculate NNTs. Noteworthy, the majority of included RCTs (13 out of 17) analysed binary outcomes. Studies with fixed times of follow-up are usually not prone to miscalculation of NNT because cumulative incidences equal simple proportions at the study end (Suissa et al. 2012). However, previous studies suggested that NNTs are miscalculated in at least half of RCTs with time to event outcomes (Hildebrandt, Vervölgyi & Bender 2009), (Suissa et al. 2012). In the present review, one out 4 RCTs with varying follow-up times applied a non-recommended method to calculate NNT (Shepherd et al. 2008). In that RCT, the effect of two doses of atorvastatin (80 mg or 10 mg daily) was tested, for the first occurrence of a major cardiovascular event (i.e. time to event outcome), in patients with coronary artery disease (CAD) and type 2 diabetes, with and without chronic kidney disease (Shepherd et al. 2008). Patients were followed for varying times (median, 4.8 years). Although Kaplan-Meier curves have been estimated, the authors used simple proportions of patients with the outcome to compute NNT (e.g. for patients with diabetes without CAD, 1/([62/441] – [57/444]) = 82) and concluded that 82 patients were needed to treat with 80 mg/day versus 10 mg/day to prevent one major cardiovascular event over 4.8 years (Shepherd et al. 2008). Using the cumulative incidences provided in Kaplan-Meier curves (12.5% for 80mg and 13.3% for 10mg), NNT would have been estimated at 125 over the same time horizon. This example illustrates how the use of simple proportions can lead to misleading values of NNT. Simple proportions should be used only if all patients are followed for the entire study period, as they equal cumulative incidences estimated by the Kaplan-Meier approach (Suissa 2015). Since follow-up times usually vary in RCTs, simple proportions are not valid estimates of cumulative incidences. In cases where follow-up is short and mostly complete, simple proportions and Kaplan-Meier incidences are almost similar (Suissa 2015).

As the present study assessed results from research published since 2006, two different methodologies were considered adequate for calculating NNT from RCTs where the outcome is time to an event (Altman & Andersen 1999), (Lubsen, Hoes & Grobbee 2000), (Mayne, Whalen & Vu 2006). More recently, however, the authors of a study comparing the risk difference approach (reciprocal of risk differences estimated by survival time methods) and the incidence difference approach (reciprocal of incidence rates differences) concluded that the methods based on incidence rates often lead to misleading

NNT estimates and recommended the use of survival time methods to estimate NNTs in RCTs with time to event outcomes (Bender et al. 2013). The incidence difference approach still can be used in the case of small baseline risks, strong treatment effects, and exponentially distributed survival times (Bender et al. 2013). Nevertheless, Girerd et al. argued that the two methods measure different things, but both are valid and provide complementary information regarding the absolute effect of an intervention, highlighting that the incidence rate approach assess person-years rather than persons (Girerd et al. 2014). This calculating method estimates the number of person-time (e.g. patient-years), not the absolute number of persons, needed to observe one less (or one more) event in the treatment group than in the control group (Bender et al. 2013), (Suissa et al. 2012), (Mayne, Whalen & Vu 2006), (Girerd et al. 2014), (Bender R 2014). This estimate is different from the "classical" person-based NNT, and therefore may be difficult to interpret (Bender R 2014). For example, 100 patient-years do not necessarily mean 100 individual patients treated over one year (or 50 patients treated for two years). A thorough explanation about person-based NNT, person-time-based NNT, and event-based NNT (for multiple recurrent outcome events) is provided elsewhere (Suissa et al. 2012), (Suissa 2013).

With regard to observational studies, one cohort study did not follow methodological recommendations (Graham et al. 2010). In that study, Kaplan-Meier curves and Cox proportional HRs for time to event, adjusted for confounding factors, with pioglitazone as reference, were used to test the effect of rosiglitazone on several cardiovascular adverse events (Graham et al. 2010). However, the authors applied unadjusted incidence rate differences to calculate NNTs, instead of using adjusted data. For example, at one year of follow-up, the NNT for a composite cardiovascular endpoint would be 92 from Kaplan-Meier curves rather than the 60 person-years obtained by the authors. Further, the authors interpreted person-years as number of persons treated over one year, which is not exactly the same. A detailed review and discussion of methods used to calculate NNTs from observational studies is provided elsewhere (Austin & Laupacis 2011), (Bender & Blettner 2002), (Bender et al. 2007).

The present study was not primarily aimed at the identification of all papers with methodological recommendations for calculating NNTs. For this reason, a systematic review of literature was not performed to identify such papers. This is a potential limitation of the study. Nevertheless, the literature used as source of evidence was probably adequate to the complexity of the assessment. The study focused the adherence of calculating methods to basic methodological recommendations, rather than to more complex methodological and

statistical issues. Therefore, estimates of NNT reported by studies that followed basic methodological recommendations are not necessarily correct. There are possibly other reasons that can still lead to biased estimates, but which could not be assessed with an acceptable effort. In addition, the magnitude of error produced in studies that did not follow basic methodological recommendations to calculate NNTs was not tested. Aside from some examples provided in the discussion, the calculation of correct NNTs was not sought for studies that did not follow recommendations. Lastly, the study was limited to the top-25 high impact factor journals in the "General and/or Internal Medicine" category. Whether the results in other fields are likely to show similar results deserves further testing.

The present results illustrate that these metrics have not been always adequately calculated. From the clinicians point of view this may rise some concerns, since these metrics can be used to support clinical decision making processes, including the prescription of medicines. Therefore, clinicians need to rely on the methodological appropriateness of such calculations.

The NNT helps to quantify the magnitude of effects of medical interventions in an absolute scale, therefore bringing added value to decisions on drug utilization for clinicians, regulators and other stakeholders. However, they should be aware that the calculation and interpretation of the NNT depend on the characteristics of a given study, namely the design and outcome variables. Moreover, they must acknowledge that a NNT is specific to a given comparison. Therefore baseline risks, clearly defined outcomes, time horizons, as well as confidence intervals should be provided. The presentation of a NNT alone, i.e. without its context, would be ambiguous and less useful for decision-making.

This study showed that, although the concept of NNT has been introduced several years ago, there are basic methodological recommendations still not being followed, particularly in meta-analyses, leading to miscalculated and misinterpreted results. Further research is needed to confirm present findings and to explore the influence of other methodological aspects that may impact the calculation of the NNT in clinical studies.

V.6. REFERENCES

ALONSO-COELLO, P. ET AL., 2016. SYSTEMATIC REVIEWS EXPERIENCE MAJOR LIMITATIONS IN REPORTING ABSOLUTE EFFECTS. J CLIN EPIDEMIOL, 72, PP. 16-26.

ALTMAN, D., 1998. CONFIDENCE INTERVALS FOR THE NUMBER NEEDED TO TREAT. BMJ, 317(7168), PP. 1309-1312.

ALTMAN, D. & ANDERSEN, P., 1999. CALCULATING THE NUMBER NEEDED TO TREAT FOR TRIALS WHERE THE OUTCOME IS TIME TO AN EVENT. BMJ, 319(7223), PP. 1492-1495.

ALTMAN, D. & DEEKS, J., 2002. META-ANALYSIS, SIMPSON'S PARADOX, AND THE NUMBER NEEDED TO TREAT. BMC MED RES METHODOL, 2(3).

ALTMAN, D. ET AL., 2011. THE REVISED CONSORT STATEMENT FOR REPORTING RANDOMIZED TRIALS: EXPLANATION AND ELABORATION. ANN INTERN MED, 134(8), PP. 663-694.

ALVES, C., BATEL-MARQUES, F. & MACEDO, A., 2012. DATA SOURCES ON DRUG SAFETY EVALUATION: A REVIEW OF RECENT PUBLISHED META-ANALYSES. PHARMACOEPIDEMIOL DRUG SAF, 21(1), PP. 21-33.

ATTIA, J., PAGE, J., HELLER, R. & DOBSON, A., 2002. IMPACT NUMBERS IN HEALTH POLICY DECISIONS. J EPIDEMIOL COMMUNITY HEALTH, 56(8), PP. 600-605.

AUSTIN, P., 2010. ABSOLUTE RISK REDUCTIONS, RELATIVE RISKS, RELATIVE RISK REDUCTIONS, AND NUMBERS NEEDED TO TREAT CAN BE OBTAINED FROM A LOGISTIC REGRESSION MODEL. J CLIN EPIDEMIOL, 63(1), PP. 2-6.

AUSTIN, P. & LAUPACIS, A., 2011. A TUTORIAL ON METHODS TO ESTIMATING CLINICALLY AND POLICY-MEANINGFUL MEASURES OF TREATMENT EFFECTS IN PROSPECTIVE OBSERVATIONAL STUDIES: A REVIEW. INT J BIOSTAT, 7(1), P. 6.

BENDER, R., KROMP, M., KIEFER, C. & STURTZ, S., 2014. ESTIMATION OF NUMBERS NEEDED TO TREAT SHOULD BE BASED ON ABSOLUTE RISKS. J CLIN EPIDEMIOL, 67(2), PP. 238-239.

BENDER, R., 2005. NUMBER NEEDED TO TREAT (NNT). IN: ENCYCLOPEDIA OF BIOSTATISTICS. CHICHESTER: WILEY, PP. 3752-3761.

BENDER, R. & BLETTNER, M., 2002. CALCULATING THE "NUMBER NEEDED TO BE EXPOSED" WITH ADJUSTMENT FOR CONFOUNDING VARIABLES IN EPIDEMIOLOGICAL STUDIES. J CLIN EPIDEMIOL, 55(5), PP. 525-530.

BENDER, R., KROMP, M., KIEFER, C. & STURTZ, S., 2013. ABSOLUTE RISKS RATHER THAN INCIDENCE RATES SHOULD BE USED TO ESTIMATE THE NUMBER NEEDED TO TREAT FROM TIME-TO-EVENT DATA. J CLIN EPIDEMIOL, 66(9), PP. 1038-1044.

BENDER, R., KUSS, O., HILDEBRANDT, M. & GEHRMANN, U., 2007. ESTIMATING ADJUSTED NNT MEASURES IN LOGISTIC REGRESSION ANALYSIS. STAT MED, 26(30), PP. 5586-5595.

BJERRE, L. & LELORIER, J., 2000. EXPRESSING THE MAGNITUDE OF ADVERSE EFFECTS IN CASE-CONTROL STUDIES: "THE NUMBER OF PATIENTS NEEDED TO BE TREATED FOR ONE ADDITIONAL PATIENT TO BE HARMED". BMJ, 320(7233), PP. 503-506.

BMJ, 2016. FOR AUTHORS, RESOURCES FOR AUTHORS, ARTICLE TYPES, RESEARCH. [ONLINE] AVAILABLE AT: HTTP://WWW.BMJ.COM/ABOUT-BMJ/RESOURCES-AUTHORS/ARTICLE-TYPES/RESEARCH [ACCESSED 13 OCT 2016].

BROWN, E., WOOD, L. & WOOD, S., 1999. THE MEDICAL DICTIONARY FOR REGULATORY ACTIVITIES (MEDDRA). DRUG SAF, 20(2), PP. 109-117.

CATES, C., 2002. SIMPSON'S PARADOX AND CALCULATION OF NUMBER NEEDED TO TREAT FROM META-ANALYSIS. BMC MED RES METHODOL, 2(1).

CITROME, L., 2010. RELATIVE VS. ABSOLUTE MEASURES OF BENEFIT AND RISK: WHAT'S THE DIFFERENCE? ACTA PSYCHIATR SCAND, 121(2), PP. 94-102.

CITROME, L. & KETTER, T., 2013. WHEN DOES A DIFFERENCE MAKE A DIFFERENCE? INTERPRETATION OF NUMBER NEEDED TO TREAT, NUMBER NEEDED TO HARM, AND LIKELIHOOD TO BE HELPED OR HARMED. INT J CLIN PRACT, 67(5), PP. 407-411.

COOK, R. & SACKETT, D., 1995. THE NUMBER NEEDED TO TREAT: A CLINICALLY USEFUL MEASURE OF TREATMENT EFFECT. BMJ, 310(6977), PP. 452-454.

DA COSTA, B. ET AL., 2012. METHODS TO CONVERT CONTINUOUS OUTCOMES INTO ODDS RATIOS OF TREATMENT RESPONSE AND NUMBERS NEEDED TO TREAT: META-EPIDEMIOLOGICAL STUDY. INT J EPIDEMIOL, 41(5), PP. 1445-1459.

DEEKS, J., HIGGINS, J. & ALTMAN, D., 2011. CHAPTER 9: ANALYSING DATA AND UNDERTAKING META-ANALYSES. IN: T. C. COLLABORATION, ED. COCHRANE HANDBOOK FOR SYSTEMATIC REVIEWS OF INTERVENTIONS VERSION 5.1.0., S.L.: THE COCHRANE COLLABORATION.

ENGELS, E. ET AL., 2000. HETEROGENEITY AND STATISTICAL SIGNIFICANCE IN META-ANALYSIS: AN EMPIRICAL STUDY OF 125 META-ANALYSES. STAT MED, 19(13), PP. 1707-1728.

FURUKAWA, T., GUYATT, G. & GRIFFITH, L., 2002. CAN WE INDIVIDUALIZE THE 'NUMBER NEEDED TO TREAT'? AN EMPIRICAL STUDY OF SUMMARY EFFECT MEASURES IN META-ANALYSES. INT J EPIDEMIOL, 31(1), PP. 72-76.

GIRERD, N., RABILLOUD, M., DUARTE, K. & ROY, P., 2014. NUMBER NEEDED TO TREAT FROM ABSOLUTE RISK AND INCIDENCE RATE: HOW TO MAKE APPLES AND ORANGES COMPARABLE? J CLIN EPIDEMIOL, 67(2), PP. 236-238.

GRAHAM, D. ET AL., 2010. RISK OF ACUTE MYOCARDIAL INFARCTION, STROKE, HEART FAILURE, AND DEATH IN ELDERLY MEDICARE PATIENTS TREATED WITH ROSIGLITAZONE OR PIOGLITAZONE. JAMA, 304(4.), PP. 411-418.

GREEN, H., PAUL, M., VIDAL, L. & LEIBOVICI, L., 2007. PROPHYLAXIS OF PNEUMOCYSTIS PNEUMONIA IN IMMUNOCOMPROMISED NON-HIV-INFECTED PATIENTS: SYSTEMATIC REVIEW AND META-ANALYSIS OF RANDOMIZED CONTROLLED TRIALS. MAYO CLIN PROC, 82(9), PP. 1052-1059.

HELLER, R. ET AL., 2003. COMMUNICATING RISKS AT THE POPULATION LEVEL: APPLICATION OF POPULATION IMPACT NUMBERS. BMJ, 327(7424), PP. 1162-1165.

HELLER, R., DOBSON, A., ATTIA, J. & PAGE, J., 2002. IMPACT NUMBERS: MEASURES OF RISK FACTOR IMPACT ON THE WHOLE POPULATION FROM CASE-CONTROL AND COHORT STUDIES. J EPIDEMIOL COMMUNITY HEALTH, 56(8), PP. 606-610.

HELLER, R., EDWARDS, R. & MCELDUFF, P., 2003. IMPLEMENTING GUIDELINES IN PRIMARY CARE: CAN POPULATION IMPACT MEASURES HELP? BMC PUBLIC HEALTH, 3(7).

HILDEBRANDT, M., VERVÖLGYI, E. & BENDER, R., 2009. CALCULATION OF NNTS IN RCTS WITH TIME-TO-EVENT OUTCOMES: A LITERATURE REVIEW. BMC MED RES METHODOL, 9(21).

HUTTON, J., 2000. NUMBER NEEDED TO TREAT: PROPERTIES AND PROBLEMS. J R STATIST SOC A, 163(3), PP. 403-419.

KLAWITER, E., CROSS, A. & NAISMITH, R., 2009. THE PRESENT EFFICACY OF MULTIPLE SCLEROSIS THERAPEUTICS: IS THE NEW 66% JUST THE OLD 33%? NEUROLOGY, 73(12), PP. 984-990.

LAUBENDER, R. & BENDER, R., 2010. ESTIMATING ADJUSTED RISK DIFFERENCE (RD) AND NUMBER NEEDED TO TREAT (NNT) MEASURES IN THE COX REGRESSION MODEL. STAT MED, 29(7-8), PP. 851-859.

LAUBENDER, R. & BENDER, R., 2014. A NOTE ON CALCULATING ASYMPTOTIC CONFIDENCE INTERVALS FOR THE ADJUSTED RISK DIFFERENCE AND NUMBER NEEDED TO TREAT IN THE COX REGRESSION MODEL. STAT MED, 33(5), PP. 798-810.

LAUPACIS, A., SACKETT, D. & ROBERTS, R., 1988. AN ASSESSMENT OF CLINICALLY USEFUL MEASURES OF THE CONSEQUENCES OF TREATMENT. N ENGL J MED, 318(26), PP. 1728-1733.

LUBSEN, J., HOES, A. & GROBBEE, D., 2000. IMPLICATIONS OF TRIAL RESULTS: THE POTENTIALLY MISLEADING NOTIONS OF NUMBER NEEDED TO TREAT AND AVERAGE DURATION OF LIFE GAINED. LANCET, 356(9243), PP. 1757-1759.

MAHER, A. ET AL., 2011. EFFICACY AND COMPARATIVE EFFECTIVENESS OF ATYPICAL ANTIPSYCHOTIC MEDICATIONS FOR OFF-LABEL USES IN ADULTS: A SYSTEMATIC REVIEW AND META-ANALYSIS. JAMA, 306(12), PP. 1359-1369.

MARX, A. & BUCHER, H., 2003. NUMBERS NEEDED TO TREAT DERIVED FROM META-ANALYSIS: A WORD OF CAUTION. ACP J CLUB, 138(2), PP. A11-A12.

MAYNE, T., WHALEN, E. & VU, A., 2006. ANNUALIZED WAS FOUND BETTER THAN ABSOLUTE RISK REDUCTION IN THE CALCULATION OF NUMBER NEEDED TO TREAT IN CHRONIC CONDITIONS. J CLIN EPIDEMIOL, 59(3), PP. 217-223.

MCALISTER, F., 2008. THE "NUMBER NEEDED TO TREAT" TURNS 20 - AND CONTINUES TO BE USED AND MISUSED. CMAJ, 179(6), PP. 549-553.

MCQUAY, H. & MOORE, R., 1997. USING NUMERICAL RESULTS FROM SYSTEMATIC REVIEWS IN CLINICAL PRACTICE. ANN INTERN MED, 126(9), PP. 712-720.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2015. NUMBER NEEDED TO HARM IN THE POST-MARKETING SAFETY EVALUATION: RESULTS FOR ROSIGLITAZONE AND PIOGLITAZONE. PHARMACOEPIDEMIOL DRUG SAF, 24(12), PP. 1259-1270.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2016A. TESTING THE USEFULNESS OF THE NUMBER NEEDED TO TREAT TO BE HARMED (NNTH) IN BENEFIT-RISK EVALUATIONS: CASE STUDY WITH MEDICINES WITHDRAWN FROM THE EUROPEAN MARKET DUE TO SAFETY REASONS. EXPERT OPIN DRUG SAF, 15(10), PP. 1301-1312.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2016B. BENEFIT-RISK OF THERAPIES FOR RELAPSING-REMITTING MULTIPLE SCLEROSIS: TESTING THE NUMBER NEEDED TO TREAT TO BENEFIT (NNTB), NUMBER NEEDED TO TREAT TO HARM (NNTH) AND THE LIKELIHOOD TO BE HELPED OR HARMED (LHH): A SYSTEMATIC REVIEW AND META-ANALYSIS. CNS DRUGS, 30(10), PP. 909-929.

MOHER, D. ET AL., 2010. CONSORT 2010 EXPLANATION AND ELABORATION: UPDATED GUIDELINES FOR REPORTING PARALLEL GROUP RANDOMISED TRIALS. BMJ, VOLUME 340, P. C869.

MOORE, R. ET AL., 2002. POOLING DATA FOR NUMBER NEEDED TO TREAT: NO PROBLEMS FOR APPLES. BMC MED RES METHODOL, VOLUME 2, P. 2.

MT-ISA, S. ET AL., 2014. BALANCING BENEFIT AND RISK OF MEDICINES: A SYSTEMATIC REVIEW AND CLASSIFICATION OF AVAILABLE METHODOLOGIES. PHARMACOEPIDEMIOL DRUG SAF, 23(7), PP. 667-678.

NUOVO, J., MELNIKOW, J. & CHANG, D., 2002. REPORTING NUMBER NEEDED TO TREAT AND ABSOLUTE RISK REDUCTION IN RANDOMIZED CONTROLLED TRIALS. JAMA, 287(21), PP. 2813-2814.

PREISS, D., ET AL. 2011. RISK OF INCIDENT DIABETES WITH INTENSIVE-DOSE COMPARED WITH MODERATE-DOSE STATIN THERAPY: A META-ANALYSIS. JAMA, 305(24), PP. 2556-2564.

SCHMID, C., LAU, J., MCINTOSH, M. & CAPPELLERI, J., 1998. AN EMPIRICAL STUDY OF THE EFFECT OF THE CONTROL RATE AS A PREDICTOR OF TREATMENT EFFICACY IN META-ANALYSIS OF CLINICAL TRIALS. STAT MED, 17(17), PP. 1923-1942.

SHEPHERD, J. ET AL., 2008. INTENSIVE LIPID LOWERING WITH ATORVASTATIN IN PATIENTS WITH CORONARY ARTERY DISEASE, DIABETES, AND CHRONIC KIDNEY DISEASE. MAYO CLIN PROC, 83(8), PP. 870-879.

SMEETH, L., HAINES, A. & EBRAHIM, S., 1999. NUMBERS NEEDED TO TREAT DERIVED FROM META-ANALYSES–SOMETIMES INFORMATIVE, USUALLY MISLEADING. BMJ, 318(7197), PP. 1548-1551.

STANG, A., POOLE, C. & BENDER, R., 2010. COMMON PROBLEMS RELATED TO THE USE OF NUMBER NEEDED TO TREAT. J CLIN EPIDEMIOL, 63(8), PP. 820-825.

STRAUS, S., GLASZIOU, P., RICHARDSON, W. & HAYNES, R., 2011. EVIDENCE-BASED MEDICINE: HOW TO PRACTICE AND TEACH IT. 4TH ED. LONDON, UK: CHURCHILL LIVINGSTONE.

SUISSA, D., BRASSARD, P., SMIECHOWSKI, B. & SUISSA, S., 2012. NUMBER NEEDED TO TREAT IS INCORRECT WITHOUT PROPER TIME-RELATED CONSIDERATIONS. J CLIN EPIDEMIOL, 65(1), PP. 42-46.

SUISSA, S., 2009. CALCULATION OF NUMBER NEEDED TO TREAT. N ENGL J MED, 361(4), PP. 424-425.

SUISSA, S., 2013. NUMBER NEEDED TO TREAT IN COPD: EXACERBATIONS VERSUS PNEUMONIAS. THORAX, 68(6), PP. 540-543.

SUISSA, S., 2015. THE NUMBER NEEDED TO TREAT: 25 YEARS OF TRIALS AND TRIBULATIONS IN CLINICAL RESEARCH. RAMBAM MAIMONIDES MED J, 6(3).

THOMSON-REUTERS, 2016. JOURNAL CITATION REPORTS. [ONLINE] AVAILABLE AT: HTTP://THOMSONREUTERS.COM/EN/PRODUCTS-SERVICES/SCHOLARLY-SCIENTIFIC-RESEARCH/RESEARCH-MANAGEMENT-AND-EVALUATION/JOURNAL-CITATION-REPORTS.HTML [ACCESSED 24 AUG 2016].

V.7. SUPPLEMENTAL DATA V

V.7.1. SUPPLEMENTAL DATA V.1 – LIST OF JOURNALS CONSIDERED IN THE LITERATURE SEARCH

Supplemental Table V. 1 – List of the 25 Journals of "General and/or Internal Medicine" with higher Impact Factor in 2015.

Rank	Full Journal Title	Total Cites	Journal Impact Factor
1	NEW ENGLAND JOURNAL OF MEDICINE	283,525	59.558
2	LANCET	195,553	44.002
3	JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	129,909	37.684
4	BMJ-BRITISH MEDICAL JOURNAL	93,118	19.697
5	ANNALS OF INTERNAL MEDICINE	49,618	16.440
6	JAMA INTERNAL MEDICINE	5,590	14.000
7	PLOS MEDICINE	20,499	13.585
8	BMC MEDICINE	7,331	8.005
9	JOURNAL OF CACHEXIA SARCOPENIA AND MUSCLE	901	7.883
10	JOURNAL OF INTERNAL MEDICINE	9,090	7.803
11	CANADIAN MEDICAL ASSOCIATION JOURNAL	12,420	6.724
12	MAYO CLINIC PROCEEDINGS	10,745	5.920
13	AMERICAN JOURNAL OF MEDICINE	22,561	5.610
14	ANNALS OF FAMILY MEDICINE	3,879	5.087
15	TRANSLATIONAL RESEARCH	2,418	4.557
16	AMERICAN JOURNAL OF PREVENTIVE MEDICINE	17,735	4.465
17	ANNALS OF MEDICINE	4,012	3.763
18	DEUTSCHES ARZTEBLATT INTERNATIONAL	2,403	3.738
19	PALLIATIVE MEDICINE	3,714	3.685
20	JOURNAL OF GENERAL INTERNAL MEDICINE	14,808	3.494
21	MEDICAL JOURNAL OF AUSTRALIA	9,739	3.369
22	AMERICAN JOURNAL OF CHINESE MEDICINE	2,535	2.959
23	BRITISH MEDICAL BULLETIN	3,727	2.921
24	PREVENTIVE MEDICINE	12,516	2.893
25	QJM-AN INTERNATIONAL JOURNAL OF MEDICINE	5,309	2.824

Source: InCitesTM Journal Citation Reports® by Thomson Reuters.

V.7.2. SUPPLEMENTAL DATA V.2 – LITERATURE SEARCH STRATEGY

Supplemental Table V. 2 – Search strategy used to identify studies reporting number needed to treat (NNT), performed in Pubmed on 24th August 2016.

Search	Terms	Results
#1	Search ((((((((((((((((((((((("The New England journal of medicine"[Journal]) OR "Lancet (London, England)"[Journal]) OR "JAMA"[Journal]) OR "British medical journal"[Journal]) OR "Annals of internal medicine"[Journal]) OR "JAMA internal medicine"[Journal]) OR "PLoS medicine"[Journal]) OR "BMC medicine"[Journal]) OR ("Journal of cachexia, sarcopenia and muscle"[Journal])) OR "Journal of internal medicine"[Journal]) OR "Canadian Medical Association journal"[Journal]) OR "Mayo Clinic proceedings"[Journal]) OR "The American journal of medicine"[Journal]) OR "Annals of family medicine"[Journal]) OR ("Translational research : the journal of laboratory and clinical medicine"[Journal])) OR "American journal of preventive medicine"[Journal]) OR "Annals of medicine"[Journal]) OR "Deutsches Ärzteblatt international"[Journal]) OR "Palliative medicine"[Journal]) OR "Journal of general internal medicine"[Journal]) OR "The Medical journal of Australia"[Journal]) OR "The American journal of Chinese medicine"[Journal]) OR "British medical bulletin"[Journal]) OR "Preventive medicine"[Journal]) OR "QJM : monthly journal of the Association of Physicians"[Journal])	560760
#2	Search numbers needed to treat[MeSH Terms]	160
#3	Search "nnt"	2333
#4	Search "nnh"	639
#5	Search "nntb"	216
#6	Search "nnth"	110
#7	Search "number needed to treat"[Title/Abstract]	3555
#8	Search "number needed to harm"[Title/Abstract]	579
#9	Search ((#2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8))	5040
#10	Search (#1 AND #9)	225
#11	Search ((#1 AND #9)) Sort by: PublicationDate Filters: Publication date from 2006/01/01 to 2015/12/31	138

V.7.3. SUPPLEMENTAL DATA V.3 – DESCRIPTION AND CHARACTERISTICS OF NNT

Supplemental Table V. 3 – List of queries used to describe and categorize NNT in selected studies.

Description and categorization of NNT estimates:
a) What was the type of variable used to compute NNT for the outcome of interest? ▪ Binary; ▪ Time-to-event.
b) Was the NNT presented together with the result of a relative effect measure? ▪ Yes; ▪ No.
c) Which was the relative effect measure presented together with the NNT? ▪ Hazard Ratio; ▪ Odds Ratio; ▪ Rate Ratio; ▪ Relative Risk; ▪ Not applicable;
d) For which study outcome was the NNT calculated? ▪ Primary outcome; ▪ Primary and secondary outcomes; ▪ Other outcomes than the primary outcome.
e) Was the NNT calculated for beneficial, harmful or both beneficial and harmful outcomes? ▪ NNTB (beneficial outcome); ▪ NNTH (harmful outcome); ▪ NNTB and NNTH.
f) Which type of NNT was calculated? ▪ Patient-based NNT; ▪ Patient-time-based NNT; ▪ Event-based NNT (multiple events).
g) Was the control event rate used to derive the NNT presented? ▪ Yes; ▪ No.
h) Was the time horizon clearly defined for the NNT? ▪ Yes; ▪ No.
i) Were the confidence intervals provided for the NNT? ▪ Yes; ▪ No.

Note: Patient-based NNT: Number of patients with outcome of interest divided by the total amount of participating patients; Patient-time-based NNT: Number of patients with outcome of interest divided by the total amount of patient-time, to account for varying follow-up times; Event-based NNT (multiple events): Number of outcome events divided by the total amount of patient-time (Suissa et al. 2012).

Supplemental Table V. 4 – List of queries used to assess methodologies used to calculate NNT in selected studies.

Assessment of the methodology used to calculate NNT:
a) Was the method used to compute NNT defined in the methods section? ▪ Yes; ▪ No; b) Which method was used to derive the NNT? ▪ Risk difference (i.e. absolute risk reduction or increase); ▪ Relative effect measure (e.g., Hazard Ratio; Odds Ratio; Rate Ratio; Relative Risk). c) What was the source of data used to calculate NNT? ▪ Simple proportions; ▪ Relative effect measure; ▪ Cumulative incidence rates (i.e. using a Kaplan-Meier approach or a Cox regression model); ▪ Pooled risk differences (i.e. derived from meta-analysis); ▪ Average risk difference; d) Was the method used to derive the NNT in line with recommendations for study design? ▪ Yes; ▪ No.

V.7.4. SUPPLEMENTAL DATA V.4 – SUPPLEMENTAL SEARCH STRATEGY TO IDENTIFY STUDIES ABOUT METHODS TO CALCULATE NNT

Supplemental Table V. 5 – Search strategy used to identify studies investigating methods for calculating number needed to treat (NNT), performed in Pubmed on 24th August 2016.

Search	Terms	Results
#1	Search numbers needed to treat[MeSH Terms]	160
#2	Search nnt	2333
#3	Search nnh	639
#4	Search nntb	216
#5	Search nnth	110
#6	Search "number needed to treat"[Title/Abstract]	3555
#7	Search "number needed to harm"	579
#8	Search (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7)	5040
#9	Search ("Epidemiologic Methods"[Majr] OR "Data Collection"[Majr] OR "Data Interpretation, Statistical"[Majr] OR "Statistics as Topic"[Majr] OR "Evidence-Based Medicine"[Mesh])	619049
#10	Search (#8 AND #9)	629

V.7.5. SUPPLEMENTAL DATA V.5 – CHARACTERISTICS OF INCLUDED STUDIES

Supplemental Table V. 6 – Main characteristics of included studies.

Reference	Journal	Year	Country	Studies, N=	Participants, N=	Study duration	Disease or condition	Disease, SOC	Intervention	Control	Primary outcome measure
Systematic review and Meta-analysis											
Palmerini et al.	Lancet	2015	Italy	10	31,666	NA	Patients with CAD undergoing PCI	Cardiac disorders	Short-duration DAPT	Long-duration (>1 yr) DAPT	All-cause mortality
Chatterjee et al.	JAMA	2014	USA	16	2,115	NA	Pulmonary embolism	Respiratory, thoracic and mediastinal disorders	Thrombolytic therapy	Conventional anticoagulant therapy	All-cause mortality and major bleeding
Jonas et al.	JAMA	2014	USA	123*	22,803	NA	Alcoholism	Psychiatric disorders	Medication for alcoholism disorders	Placebo or other medication	Not clearly defined. Several outcomes analysed: alcohol consumption, motor vehicle crashes, injuries, quality of life, function, mortality, and harms
Spielmans et al.	PLoS Med	2013	USA	7	3,549	NA	Depression	Psychiatric disorders	Atypical antipsychotics	Placebo	A primary outcome is not clearly defined. Several outcomes analysed: remission, response, and adverse events.
Kayentao et al.	JAMA	2013	UK	7	6,281	NA	Pregnancy	Pregnancy, puerperium and perinatal conditions	3 or more doses of sulfadoxine-pyrimethamine	2 doses of sulfadoxine-pyrimethamine	Low birth weight (LBW); and mean birth weight
Hempel et al.	JAMA	2012	USA	82	11,811	NA	Antibiotic-associated diarrhea	Gastrointestinal disorders	Probiotics (Lactobacillus, Bifidobacterium, Saccharomyces, Streptococcus, Enterococcus, and/orBacillus)	No treatment, placebo, or a different probiotic or probiotic dose	Participants with diarrhea
Leucht et al.	Lancet	2012	Germany	65	6,493	NA	Schizophrenia	Psychiatric disorders	Antipsychotics	Placebo	Relapse between 7 and 12 months
Shah et al.	Am J Med	2012	USA	18	8,595	NA	Irritable bowel syndrome	Gastrointestinal disorders	Alosetron; tricyclic antidepressants; rifaximin; lubiprostone; or selective serotonin reuptake inhibitors	Placebo	Adverse event requiring discontinuation of treatment
Maher et al.	JAMA	2011	USA	393		NA	Off-label conditions (agitation in dementia, anxiety, and OCD)	Injury, poisoning and procedural complications	Atypical antipsychotics (risperidone, olanzapine, quetiapine, aripiprazole, ziprasidone, asenapine, iloperidone, or paliperidone)	Placebo, another atypical antipsychotic medication, or other pharmacotherapy for adult off-label conditions	A primary outcome is not clearly defined. Several outcomes analysed: improvement in psychosis, improvement in agitation, and a total global score.
Preiss et al.	JAMA	2011	UK	5	32,752	NA	Cardiovascular disease	Cardiac disorders	Intensive-dose statin therapy	Moderate-dose statin therapy	Major cardiovascular events (prevention); New-onset diabetes (adverse event)

Reference	Journal	Year	Country	Studies, N=	Participants, N=	Study duration	Disease or condition	Disease, SOC	Intervention	Control	Primary outcome measure
Shamliyan et al.	J Gen Intern Med	2011	USA	16	4,431	NA	Chronic hepatitis B	Infections and infestations	Antiviral drugs	Antiviral drugs or placebo	A primary outcome is not clearly defined. Several outcomes analysed: mortality, incidence of hepatocellular carcinoma, liver failure, prevalence and incidence of cirrhosis, HBeAg or HBsAg presence or seroconversion, viral load of HBV deoxyribonucleotide acid (HBV DNA), ALT levels, histological necroinflammatory and fibrosis scores, and adverse events.
Coker et al.	JAMA	2010	USA	7	2,058	NA	Acute Otitis Media	Infections and infestations	Antibiotics (immediate initiation)	Placebo or strategy of observation with possible delayed treatment	Clinical success
Testa et al.	Q J Med	2008	UK	8	1,318	NA	Myocardial infarction	Cardiac disorders	Rescue PCI or repeat thrombolysis	Conservative therapy	A primary outcome is not clearly defined. Several outcomes analysed: major adverse events, defined as the composite of overall mortality and re-infarction; stroke, congestive heart failure (CHF), major bleeds, and minor bleeds
Bangalore et al.	Lancet	2008	USA	33	12,306	NA	Non-cardiac surgery	Surgical and medical procedures	β blockers (intravenous or oral)	Other drugs, placebo, or no intervention	A primary outcome is not clearly defined. Several outcomes analysed: 30-day all-cause mortality, cardiovascular mortality, non-fatal myocardial infarction, non-fatal stroke, and heart failure; perioperative adverse events (bradycardia, hypotension, and bronchospasm)
Christensen et al.	Lancet	2007	Denmark	4	4,105	NA	Obesity	Metabolism and nutrition disorders	Rimonabant 20 mg	Placebo	Difference in mean weight change and the number of individuals achieving at least 10% weight reduction handled as a dichotomous responder criterion
Green et al.	Mayo Clin Proc	2007	Israel	12	1,245	NA	Pneumocystis pneumonia (PCP) prophylaxis (caused by Pneumocystis jirovecii)	Infections and infestations	Antibiotics	Placebo, no intervention, or antibiotics with no activity against P jirovecii	PCP infection
Bridge et al.	JAMA	2007	USA	27	5,31	NA	Major depressive disorder (MDD), OCD, and non-OCD anxiety disorders	Psychiatric disorders	SSRIs and other second generation antidepressants	Placebo	Efficacy: treatment response and the prospectively identified scalar variable assessing change in symptoms from baseline to the end of treatment; safety: suicidal ideation/suicide attempt
Leontiadis et al.	Mayo Clin Proc	2007	USA	24	4,373	NA	Peptic ulcer bleeding	Gastrointestinal disorders	Proton pump inhibitors (PPIs)	Placebo or a histamine 2-receptor antagonist	Mortality from any cause within 30 days of randomization
Dentali et al.	Ann Intern Med	2007	Canada	9	19,958	NA	Hospitalized medical patients at risk for venous thromboembolism	Vascular disorders	Anticoagulant prophylaxis	No anticoagulant prophylaxis	A primary outcome is not clearly defined. Several outcomes analysed: any pulmonary embolism (PE), fatal PE, symptomatic deep venous thrombosis (DVT), and all-cause mortality; and major bleeding
Rovers et al.	Lancet	2006	The Netherlands	6	1,643	NA	Acute media otitis	Infections and infestations	Antibiotics (immediate initiation)	Placebo (or delayed treatment with antibiotics)	Extended course of acute otitis media, consisting of pain, fever, or both at 3–7 days

Reference	Journal	Year	Country	Studies, N=	Participants, N=	Study duration	Disease or condition	Disease, SOC	Intervention	Control	Primary outcome measure
Hollingsworth et al.	Lancet	2006	USA	9	693	NA	Urinary stone disease	Renal and urinary disorders	Calcium-channel blockers or α -blockers	Non-use of calcium-channel blockers or α -blockers	Proportion of patients who passed stones (cumulative incidence)
Bongartz et al.	JAMA	2006	UK	9	41,005	NA	Rheumatoid arthritis	Musculoskeletal and connective tissue disorders	Anti-TNF therapy	Placebo	Serious infections and malignancies
Spiegel et al.	Am J Med	2006	USA	26	41,529	NA	Chronic arthritis pain	Musculoskeletal and connective tissue disorders	COX-2 inhibitors or NSAIDs + PPIs	NSAIDs	Dyspepsia
Randomized Controlled Trial											
Lenze et al.	Lancet	2015	Canada	NA	181	12 weeks	Depression	Psychiatric disorders	Aripiprazole	Placebo	Remission of depression
Unger et al.	BMC Medicine	2015	Australia	NA	2,793	Until delivery	Pregnancy	Pregnancy, puerperium and perinatal conditions	Sulphadoxine-pyrimethamine plus azithromycin	Sulphadoxine-pyrimethamine and chloroquine plus placebo	Proportion of live born, singleton infants without congenital malformations with low birthweight
Imazio et al.	JAMA	2014	Italy	NA	360	3 months	Patients undergoing cardiac surgery	Surgical and medical procedures	Colchicine	Placebo	Occurrence of postpericardiotomy syndrome
Imazio et al.	Lancet	2014	Italy	NA	240	18 months	Recurrent pericarditis	Cardiac disorders	Colchicine	Placebo	Recurrence of pericarditis
Lazzerini et al.	JAMA	2013	Italy	NA	54	8 weeks	Refractory paediatric Crohn's disease	Gastrointestinal disorders	Thalidomide	Placebo	Clinical remission; Reduction in Pediatric Crohn Disease Activity Index (PCDAI) score by 25% or 75%
Mason et al.	JAMA Intern Med	2014	USA	NA	150	12 weeks	Alcoholism	Psychiatric disorders	Gabapentin	Placebo	Complete abstinence; no heavy drinking
Liou et al.	Lancet	2013	Taiwan	NA	900	7 weeks	H pylori infection	Infections and infestations	S14	T14	Eradication rate in first-line treatment
Enden et al.	Lancet	2012	Norway	NA	209	24 months	Deep vein thrombosis	Vascular disorders	Catheter-directed thrombolysis (CDT) using alteplase + Conventional treatment	Conventional treatment with initial low molecular weight heparin (LMWH) and warfarin followed by warfarin alone	Iliofemoral patency after 6 months; and frequency of post-thrombotic syndrome (PTS) after 24 months
Ryan et al.	Lancet	2012	Australia	NA	62	8 weeks	Refractory chronic cough	Respiratory, thoracic and mediastinal disorders	Gabapentin	Placebo	Change in cough-specifi c quality of life (Leicester cough questionnaire [LCQ] score) from baseline to 8 weeks of treatment
Srinivasan et al.	BMC Medicine	2012	Uganda	NA	352	7 days‡	Pneumonia	Infections and infestations	Zinc	Placebo	Time taken for normalization of respiratory rate, time taken for normalization of temperature and time taken for oxygen saturation to normalize (92% or more), while breathing room air
Franklin et al.	JAMA	2011	USA	NA	124	12 weeks	Pediatric OCD	Psychiatric disorders	Medication management plus instructions in cognitive behaviour therapy (CBT); medication management plus CBT	Medication management only	Proportion of patients responding positively to treatment by improving their baseline obsessive-compulsive scale score by 30% or more

Reference	Journal	Year	Country	Studies, N=	Participants, N=	Study duration	Disease or condition	Disease, SOC	Intervention	Control	Primary outcome measure
Imazio et al.	Ann Intern Med	2011	Italy	NA	120	18 months	Recurrent pericarditis	Cardiac disorders	Colchicine	Placebo	Recurrence rate of pericarditis
Brinks et al.	Ann Fam Med	2011	The Netherlands	NA	120	12 months	Greater trochanteric pain syndrome (GTPS)	Musculoskeletal and connective tissue disorders	Usual care and local corticosteroid injection	Usual care	Recovery at 3 and 12 months as measured on a 7-point Likert Scalre, and severity of pain during last week measured with a numeric rating scale
Zinman et al.	Lancet	2010	Canada	NA	207	3.9 years	Impaired glucose tolerance	Metabolism and nutrition disorders	Rosiglitazone + Metformin	Placebo	Development of new onset type 2 diabetes
Kenyon et al.	Lancet	2008	UK	NA	3,196	7 years	Children at age 7 years born to the women who had completed the ORACLE II Study	Pregnancy, puerperium and perinatal conditions	Erythromycin and/or amoxicillin–clavulanate	Placebo	Any level of functional impairment
Shepherd et al.	Mayo Clin Proc	2008	UK	NA	10,001	4.8 years	CAD, Diabetes, and Chronic Kidney Disease	Cardiac disorders	High-dose atorvastatin	Low-dose atorvastatin	Major cardiovascular events
Halonen et al.	JAMA	2007	Finland	NA	241	84 hours	Cardiac surgery (prevention of atrial fibrillation)	Surgical and medical procedures	Hydrocortisone	Placebo	Atrial fibrillation
Retrospective Cohort Study											
Jørgensen et al.	JAMA Intern Med	2015	Denmark	NA	55,32	30 days	Hypertension	Vascular disorders	Beta-blocker	Other antihypertensive drugs	MACE and all-cause mortality
Smith et al.	Am J Med	2015	USA	NA	11,18	1 year	Myocardial infarction	Cardiac disorders	Statin	Non-use of statin	All-cause mortality and CV hospitalizations; and AEs (diabetes mellitus and myopathy)
Parekh et al.	JAMA Intern Med	2014	USA	NA	30,411	14 days	Diabetes and concomitant infection	Metabolism and nutrition disorders	Antimicrobials known to cause hypoglycaemia	Antimicrobials not linked to hypoglycemia	Any hospitalization or emergency department visit owing to hypoglycemia within 14 days of antimicrobial exposure
Fexer et al.	Dtsch Arztebl Int	2014	Germany	NA	2,992	42 months	COPD	Respiratory, thoracic and mediastinal disorders	Teophyline	Non-use of theophyline	Hospitalizations and disease exacerbations
London et al.	JAMA	2013	USA	NA	55,138	30 days	Noncardiac surgery	Surgical and medical procedures	Beta-blocker	Non-use of beta-blocker	All-cause mortality
Meropol et al.	Ann Fam Med	2013	USA	NA	814,283	15 days	Acute nonspecific respiratory infections	Infections and infestations	Antibiotics	Non-use of antibiotics	Hospitalization within 15 days for severe adverse drug events and community-acquired pneumonia
Leung et al.	Am J Med	2011	USA	NA	2,613	10 years†	Hypertension	Vascular disorders	Thiazide	Non-use of thiazide	First occurrence of hyponatremia
Graham et al.	JAMA	2010	USA	NA	227,571	Until the earliest occurrence of a study end point	Type 2 diabetes mellitus	Metabolism and nutrition disorders	Rosiglitazone	Pioglitazone	Individual end points of acute myocardial infarction (AMI), stroke, heart failure, and all-cause death, and composite end point of AMI, stroke, heart failure, or death
Wijeysundera et al.	Lancet	2008	Canada	NA	259,037	30 days	Elective surgical procedures	Surgical and medical procedures	Epidural anaesthesia	No epidural anaesthesia	All-cause death within 30 days after surgery

Reference	Journal	Year	Country	Studies, N=	Participants, N=	Study duration	Disease or condition	Disease, SOC	Intervention	Control	Primary outcome measure
Nested case-control											
Etminan et al.	JAMA	2012	Canada	NA	989,591	1,7 years	Ophtalmological condition	Eye disorders	Oral fluoroquinolones	Non-use of fluoroquinolones	Retinal detachment
Bell et al.	JAMA	2009	Canada	NA	96,128	14 days	Cataract	Eye disorders	Tansulosin or other alfa-blocking drugs	No exposure in the year prior to cataract surgery	Postoperative ophthalmic adverse events (a composite of procedures signifying retinal detachment, lost lens or lens fragment, or endophthalmitis occurring within 14 days after cataract surgery)

AE, adverse event; AMI, acute myocardial infarction; CAD, coronary artery disease; CER, control event rate; CHF, congestive heart failure; COPD, Chronic Obstructive Pulmonary Disease; CV, cardiovascular; DAPT, dual antiplatelet therapy; DVT, deep venous thrombosis; ICU: Intensive Care Unit; MACE, major adverse cardiovascular event; mo, months; MedDRA Medical Dictionary for Regulatory Activities; NA, not applicable; NNT, number needed to treat; NNTB, number needed to treat to benefit; NNTH, number needed to treat to harm; NSAIDs, non-steroidal anti-inflammatory drugs; OCD, obsessive-compulsive disorder; PCI, percutaneous coronary intervention; PCP, Pneumocystis pneumonia; PE, pulmonary embolism; PPIs, proton-pump inhibitors; RD, risk difference; SOC, system organ class; SSRIs, selective serotonin reuptake inhibitors; TNF, tumour necrosis factor; wk, weeks; yr, years;

*122 RCT and 1 cohort study;

‡ Until hospital discharge, death or a maximum of 7 days, whichever came first;

† Subjects were followed from their index date until first occurrence of hyponatremia, death, or December 31, 2009 (whichever came first), providing for a maximum follow-up of 10 years;

¥ Extension (follow-up) of a RCT; ORACLE II: The ORACLE II trial compared the use of erythromycin and/or amoxicillin–clavulanate (co-amoxiclav) with that of placebo for women in spontaneous preterm labour and intact membranes, without overt signs of clinical infection, by use of a factorial randomised design;

₹ Median follow-up before co-trimoxazole was 154 (IQR 147–161) days, during co-trimoxazole was 532 (488–542) days, and during ART and co-trimoxazole was 749 (699–812) days.

V.7.6. SUPPLEMENTAL DATA V.6 – JOURNALS WITH STUDIES REPORTING NNT

Supplemental Table V. 7 – Number of publications reporting number needed to treat (NNT) values, according to study design and journal.

Journal	Study Design				Total
	Systematic review and meta-analysis	RCT	Cohort	Case-control	
Am J Med	2	0	2	0	4
Ann Fam Med	0	1	1	0	2
Ann Intern Med	1	1	0	0	2
BMC Medicine	0	2	0	0	2
Dtsch Arztebl Int	0	0	1	0	1
J Gen Intern Med	1	0	0	0	1
JAMA	9	4	2	2	17
JAMA Intern Med	0	1	2	0	3
Lancet	6	7	1	0	14
Mayo Clin Proc	2	1	0	0	3
PLoS Med	1	0	0	0	1
QJM	1	0	0	0	1
Total	23	17	9	2	51

RCT, Randomized controlled trial.

V.7.7. SUPPLEMENTAL DATA V.7 – ASSESSMENT OF METHODS USED TO CALCULATE NNT

Supplemental Table V. 8 – Description of data used to assess the completeness of information and the appropriateness of methods used to compute NNTs in the included studies.

Reference	Primary outcome type	Type of variable	Relative effect measure	Type of relative effect measure	NNT calculated for outcome	NNTB and/or NNTH	Type of NNT	CER presented	Time horizon defined for NNT	Confidence intervals for NNT	Methodology used to compute NNT defined in methods section	Method used to compute NNT	Source of data used to compute NNT	Adequate method to compute NNT	Comments
Systematic review and Meta-analysis															
Palmerini et al.	Efficacy	Time to event	Yes	Hazard Ratio	Primary and secondary outcomes	NNTB	Patient-time-based NNT	No	Yes	No	Yes	Relative effect measure	Relative effect measure	Yes	NNT calculated using 'metannt' command with STATA. The RD is calculated based on an assumed value of the risk in the control group. The 'metannt' command calculates this by deriving an estimate of the intervention effect (e.g. a risk ratio), applying it to a population with a given outcome event risk, and deriving from this a projected event risk if the population were to receive the intervention.
Chatterjee et al.	Efficacy and Safety	Binary	Yes	Odds ratio	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	Yes	No	Yes	Yes	Relative effect measure	Relative effect measure	Yes*	*The method used to compute NNT is adequate. However, the CER was obtained using the total number of events and total patients from all studies in the meta-analysis (Simpsons' paradox).
Jonas et al.	Efficacy and Safety	Binary	No	NA	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	No	No	Yes	Yes	NNT=1/RD	Pooled RD	No	NNTB and NNTH were calculated only when pooled RDs found a statistically significant result. A pooled RD was calculated for two outcomes. Duration of included trials ranged from 12 to 52 weeks for the outcome any drinking, and from 12 to 24 weeks for heaving drinking.
Spielmans et al.	Efficacy and Safety	Binary	Yes	Odds ratio	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	No	No	Yes	Yes	Relative effect measure	Relative effect measure	Yes	Although not presented, authors stated that "The baseline risk was calculated separately for each drug, so that placebo participants in one drug's trials were not used to calculate baseline risk for a different drug". Conversions from OR to NNT were performed using Visual Rx software.
Kayentao et al.	Efficacy	Binary	Yes	Relative risk	Primary outcome	NNTB	Patient-based NNT	Yes	Yes	No	Yes	Relative effect measure	Relative effect measure	Yes	Time horizon considered well defined because the outcome is assessed at delivery (after pregnancy).
Hempel et al.	Efficacy	Binary	Yes	Relative risk	Primary outcome	NNTB	Patient-based NNT	No	No	Yes	Yes	NNT=1/RD	Pooled RD	No	The pooled risk difference (obtained from meta-analysis) lead to a loss of follow-up time (indeed "Most trials either did not specify the follow-up period, or the assessment was explicitly limited to the time of antibiotics treatment")
Leucht et al.	Efficacy	Binary	Yes	Relative risk	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	Yes	Yes	Yes	Yes	NNT=1/RD	Pooled RD	No	The outcome is assessed between 7 and 12 months of follow-up; a mean study duration is indicated for each outcome with NNT (calculated from absolute RD pooled from the meta-analysis).

Reference	Primary outcome type	Type of variable	Relative effect measure	Type of relative effect measure	NNT calculated for outcome	NNTB and/or NNTH	Type of NNT	CER presented	Time horizon defined for NNT	Confidence intervals for NNT	Methodology used to compute NNT defined in methods section	Method used to compute NNT	Source of data used to compute NNT	Adequate method to compute NNT	Comments
Shah et al.	Safety	Binary	Yes	Relative Risk	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	No	No	Yes	Yes	NNT=1/RD	Pooled RD	No	The study comprehends the calculation and comparison of NNT for several treatments. However, NNTs are not comparable because they were calculated from pooled RDs and times of follow-up vary considerably across studies included in the meta-analysis (10 days to 48 weeks).
Maher et al.	Efficacy and Safety	Binary	Yes	Relative risk; Odds Ratio	Primary and secondary outcomes	NNTH	Patient-based NNT	No	No	No	Yes	Relative effect measure	Relative effect measure	Yes*	NNT was calculated for significant RRs (NNTB) or ORs (NNTH) using pooled RR and the assumed control risk from the placebo group). *The method is adequate; however, the authors did not clearly define the time-horizon and did not provide the control event rate used to compute NNTs.
Preiss et al.	Efficacy and Safety	Binary	Yes	Odds ratio	Primary outcome	NNTB; NNTH	Patient-time-based NNT	Yes	Yes	No	No	NNT=1/RD	Pooled RD	No	The variable for the primary outcome of the study is binary and pooled OR (95% CI) was calculated. However, NNT was calculated by taking the reciprocal of RD between pooled event rates per 1000 patient-years. Person-time based NNT was presented and interpreted as the number of persons needed to treat over one year.
Shamliyan et al.	Efficacy	Binary	No	NA	Primary and secondary outcomes	NNTB	Patient-based NNT	Yes	No	Yes	Yes	NNT=1/RD	Pooled RD	No	Several antiviral treatments were compared based on estimates NNT. However, studies with different times of follow-up for antiviral treatments were used to pool absolute RD. The time horizon factor is lost.
Coker et al.	Efficacy	Binary	No	NA	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	Yes	Yes	Yes	No	NNT=1/RD	Pooled RD	No	The pooled RD was obtained for a 14 day follow-up duration in every studies included in the meta-analysis. However, RD varies considerably across the studies included in the meta-analysis (ranging from -8% to 27%).
Testa et al.	Efficacy and Safety	Binary	Yes	Odds ratio	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	No	No	Yes	Yes	NNT=1/RD	Pooled RD	No	Pooled RD was used to calculate NNT. The follow-up of included studies ranged from 'in hospital' to 6 months.
Bangalore et al.	Efficacy and Safety	Binary	Yes	Odds ratio	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	No	Yes	No	Yes	Relative effect measure	Relative effect measure	Yes	Time-horizon was 30 days within the surgery. CERs used to compute NNTs were not presented. Although the methodology used to compute NNT was cited in the methods section, the formula was not provided.
Christensen et al.	Efficacy	Binary	Yes	Odds ratio	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	Yes	Yes	Yes	Yes	Relative effect measure	Relative effect measure	Yes	On the basis of combined OR values, NNTB and NNTH applying the overall event rate in the placebo group as a proxy for baseline risk.
Green et al.	Efficacy	Binary	Yes	Relative Risk	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	Yes	No	Yes*	Yes	Relative effect measure	Relative effect measure	Yes	*Confidence interval for the NNT was provided for the primary outcome within the primary analysis. Other NNTs were estimated according to different CERs, but without confidence intervals.
Bridge et al.	Efficacy and Safety	Binary	Yes	Relative Risk	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	Yes	No	Yes	Yes	NNT=1/RD	Pooled RD	No	DerSimonian and Laird random-effects model was used to obtain a pooled estimate of the RD (95% CI). NNT was calculated as the reciprocal of RD. The duration of follow-up and the baseline risk varied considerably across included studies.

Reference	Primary outcome type	Type of variable	Relative effect measure	Type of relative effect measure	NNT calculated for outcome	NNTB and/or NNTH	Type of NNT	CER presented	Time horizon defined for NNT	Confidence intervals for NNT	Methodology used to compute NNT defined in methods section	Method used to compute NNT	Source of data used to compute NNT	Adequate method to compute NNT	Comments
Leontiadis et al.	Efficacy	Binary	Yes	Odds ratio	Primary and secondary outcomes	NNTB	Patient-based NNT	Yes	Yes	Yes	No	Relative effect measure	Relative effect measure	Yes	NNTs were estimated using OR and unweighted pooled rates in control groups. The methodology used to compute NNT is considered correct, but it is not clearly described in methods section.
Dentali et al.	Efficacy and Safety	Binary	Yes	Relative Risk	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	Yes	No	No	Yes	NNT=1/RD	Simple proportions	No	Raw totals of patients from each study were added together to estimate proportions and calculate RD, i.e. treating data as it all came from one study (Simpson's paradox). Further, the baseline risk ranged considerably across included studies (e.g. 0.2% to 4.0% for pulmonary embolism).
Rovers et al.	Efficacy	Binary	Yes	Relative Risk	Primary and secondary outcomes	NNTB	Patient-based NNT	Yes	Yes	No	No	NNT=1/RD	Pooled RD	No	Although it is not clearly stated in the methods section, the discussion of the study suggest that the authors calculated pooled RD by means of the meta-analysis.
Hollingsworth et al.	Efficacy	Binary	Yes	Relative Risk	Primary outcome	NNTB	Patient-based NNT	Yes	No	Yes	Yes	Relative effect measure	Relative effect measure	Yes	Several CERs were indicated and used to calculate several NNT from risk ratios (as a relative risk) according to different baseline risks. CERs were not associated with a time-horizon.
Bongartz et al.	Safety	Binary	Yes	Odds ratio	Primary outcome	NNTH	Patient-based NNT	No	Yes	Yes	Yes	NNT=1/RD	Pooled RD	No	NNT calculated for treatment periods of 6 to 12 months and 3 to 12 months, using Mantel-Haenszel fixed-estimate of absolute RD in cases in which an OR of at least 1.5 was detected
Spiegel et al.	Efficacy	Binary	Yes	Relative Risk	Primary outcome	NNTB	Patient-based NNT	No	No	No	Yes	NNT=1/RD	Pooled RD	No	A pooled RD was calculated for two comparisons. Duration of included trials ranged from 6 to 78 weeks for one comparison; and from 12 to 24 weeks for another comparison.
Randomized Controlled Trial															
Lenze et al.	Efficacy	Binary	Yes	Odds ratio	Primary outcome	NNTB	Patient-based NNT	Yes	Yes	Yes	No	NNT=1/RD	Simple proportions	Yes	All patients completed the follow-up period.
Unger et al.	Efficacy	Binary	Yes	Relative Risk	Primary outcome	NNTB	Patient-based NNT	Yes	Yes	No	No	NNT=1/RD	Simple proportions	Yes	Only women with birth outcome was considered for the analysis, so there is no loss of follow-up.
Imazio et al.	Efficacy	Binary	No	NA	Primary and secondary outcomes	NNTB; NNTH	Patient-based NNT	Yes	Yes	No	No	NNT=1/RD	Simple proportions	Yes	No patients were lost to follow-up and all were analyzed for outcomes. Absolute differences were provided, but not relative effect measures.
Imazio et al.	Efficacy	Binary	Yes	Relative risk	Primary outcome	NNTB	Patient-based NNT	Yes	Yes	No	No	NNT=1/RD	Simple proportions	Yes	No patients were lost to follow-up.
Lazzerini et al.	Efficacy	Binary	Yes	Relative Risk	Primary outcome	NNTB	Patient-based NNT	Yes	Yes	Yes	No	NNT=1/RD	Simple proportions	Yes*	*Assuming that no patient was lost to follow-up since all randomized patients were considered in all calculations.
Mason et al.	Efficacy	Binary	Yes	Odds ratio	Primary outcome	NNTB	Patient-based NNT	Yes	Yes	Yes	No	NNT=1/RD	Simple proportions	Yes	Although only 57% of patients completed the study (85/150), we considered the method appropriate for computing NNT results since the authors of the study managed to classify all patients as responders or non-responders, including those that left the study earlier.

Reference	Primary outcome type	Type of variable	Relative effect measure	Type of relative effect measure	NNT calculated for outcome	NNTB and/or NNTH	Type of NNT	CER presented	Time horizon defined for NNT	Confidence intervals for NNT	Methodology used to compute NNT defined in methods section	Method used to compute NNT	Source of data used to compute NNT	Adequate method to compute NNT	Comments
Liou et al.	Efficacy	Binary	No	NA	Primary outcome	NNTB	Patient-based NNT	Yes	Yes	Yes	No	NNT=1/RD	Simple proportions	Yes	About 5% loss at follow-up; patients who did not return for a follow-up 13C-UBT were recorded as treatment failure.
Enden et al.	Efficacy	Binary	No	NA	Primary outcome	NNTB	Patient-based NNT	Yes	Yes	Yes	No	NNT=1/RD	Simple proportions	Yes	Only patients completing follow-up were assessed.
Ryan et al.	Efficacy	Binary	No	NA	Primary outcome	NNTB	Patient-based NNT	Yes	Yes	No	No	NNT=1/RD	Simple proportions	Yes	Only patients completing follow-up were assessed. Primary outcome was assessed after dichotomization of a continuous variable (mean change).
Srinivasan et al.	Efficacy	Binary	Yes	Relative risk	Secondary outcome	NNTB	Patient-based NNT	Yes	Yes	No	No	NNT=1/RD	Simple proportions	Yes	All participants completed the follow-up. However the follow-up for those dying before the 7 days was shorter.
Franklin et al.	Efficacy	Binary	No	NA	Primary outcome	NNTB	Patient-based NNT	Yes	Yes	No	No	NNT=1/RD	Simple proportions	Yes*	A total of 81.5% of patients completed the 12 weeks of follow-up. *As part of the study design, efforts were made to collect all outcomes on all randomized participants even when treatment was prematurely terminated. Prior to analysis, multiple imputation was used to replace missing values. A sequential regression multivariate imputation algorithm was used, as implemented in the IVEware package for SAS.
Imazio et al.	Efficacy	Time to event	Yes	Relative risk (reduction)	Primary outcome	NNTB	Patient-time-based NNT	Yes	Yes	Yes	No	NNT=1/RD	Cumulative incidence rates	Yes	No patient was lost to follow-up, and all participants were analyzed for outcomes according to their original assigned groups. The recurrence rate was calculated using Cox regression analysis.
Brinks et al.	Efficacy	Binary	Yes	Odds ratio	Primary outcome	NNTB	Patient-based NNT	Yes	Yes	No	No	NNT=1/RD	Simple proportions	Yes	NNT for 3 months of follow-up. Only one patient was lost to follow-up at 3 months.
Zinman et al.	Efficacy	Time to event	Yes	Hazard ratio	Primary outcome	NNTB	Patient-time-based NNT	Yes	Yes	Yes	No	NNT=1/RD	Cumulative incidence rates	Yes	Separate product-limit estimated cumulative incidence curves were calculated for the two treatment groups and compared with the log-rank test. Cox proportional hazards models were used to assess the effect of rosiglitazone and metformin on the hazard of the primary outcome.
Kenyon et al.	Safety	Binary	Yes	Odds ratio	Secondary outcome	NNTH	Patient-based NNT	Yes	Yes	Yes	No	NNT=1/RD	Simple proportions	Yes	NNT calculated for a secondary outcome from proportion of children with cerebral palsy whose mothers had participated in ORACLE II Study
Shepherd et al.	Efficacy	Time to event	Yes	Hazard ratio	Primary outcome	NNTB	Patient-time-based NNT	Yes	Yes	No	No	NNT=1/RD	Simple proportions	No	NNT calculated as 1/RD using final rates of event and citing a median time of follow-up of 4.8 years (NNT=14 in patients with diabetes and Chronic Kidney Disease). However, a Kaplan-Meier curve is provided in the study, which should have been used (since the median follow-up is lower than the 5-years objective, at least some patients did not complete the follow-up). From the Kaplan-Meier curve, we would have 20.3% and 14.0% patients with the outcome in the atorvastatin 10 mg and 80 mg/day, respectively, at 4.8 years of follow-up and a NNT = 15.8).

Reference	Primary outcome type	Type of variable	Relative effect measure	Type of relative effect measure	NNT calculated for outcome	NNTB and/or NNTH	Type of NNT	CER presented	Time horizon defined for NNT	Confidence intervals for NNT	Methodology used to compute NNT defined in methods section	Method used to compute NNT	Source of data used to compute NNT	Adequate method to compute NNT	Comments
Halonen et al.	Efficacy	Time to event	Yes	Hazard ratio	Primary outcome	NNTB	Patient-time-based NNT	Yes	Yes	No	No	NNT=1/RD	Cumulative incidence rates	Yes	Kaplan Meier curves were depicted for non-atrial fibrillation (outcome of interest). A multivariable Cox proportional hazards regression analysis model was performed to adjust for several variables.
Retrospective Cohort Study															
Jørgensen et al.	Safety	Binary	Yes	Odds ratio	Primary outcome	NNTH	Patient-based NNT	Yes	Yes	Yes	Yes	NNT=1/RD	Average RD	Yes	NNTH adjusted for all variables using the method and SAS macro by Bender and Vervolgyi, i.e. multiple logistic regression (average risk difference approach).
Smith et al.	Efficacy and Safety	Time to event	Yes	Hazard ratio	Primary outcome*	NNTB	Patient-time-based NNT	Yes	Yes	Yes**	Yes	NNT=1/RD	Cumulative incidence rates	Yes	The authors used a Cox proportional hazards regression with propensity score matching and adjustment to control for confounding. RD={[Sc(t)]^h–Sc(t)}, where Sc is the survival in control group, t is time, h is the hazard ratio. *NNT calculated only for prevention of death and CV hospitalizations. **Confidence intervals were presented only for statistically significant results.
Parekh et al.	Safety	Binary	Yes	Odds ratio	Primary outcome	NNTH	Patient-based NNT	No	Yes	Yes	Yes	NNT=1/RD	Average RD	Yes	Patients filling a prescription for one antimicrobial agent were considered. Method from Austin 2006 was used to compute NNT.
Fexer et al.	Safety	Time to event	Yes	Hazard ratio	Primary outcome	NNTH	Patient-time-based NNT	Yes	Yes	Yes	Yes	NNT=1/RD	Cumulative incidence rates	Yes	Observational study using propensity score matching. The 3.5-year observation period must be considered when interpreting NNTH results. A Kaplan–Meier estimation was used to calculate probability of exacerbation for the control and the intervention group.
London et al.	Efficacy	Binary	Yes	Relative risk	Primary and secondary outcomes	NNTB	Patient-based NNT	No	Yes	Yes	Yes	NNT=1/RD	Average RD	Yes	NNT were calculated as the inverse of the absolute risk reduction estimated in the propensity-score matched sample.
Meropol et al.	Efficacy	Binary	No	NA	Primary outcome	NNTB	Patient-based NNT	No	Yes	No	No	NNT=1/RD	Average RD	Yes	Efficacy: prevention of pneumonia. A regression model was used to calculate incidence rates and average risk differences adjusted according to several covariates.
Leung et al.	Safety	Time to event	Yes	Rate ratio	Primary and secondary outcomes	NNTH	Patient-time-based NNT	Yes	Yes	Yes	Yes	NNT=1/RD	Cumulative incidence rates	Yes	Incidence rates and incidence rate ratios were calculated using Poisson regression (adjusted for several covariates). A "survival" type curve (for cumulative incidences) is provided in the paper.
Graham et al.	Safety	Time to event	Yes	Hazard ratio	Primary outcome	NNTH	Patient-time-based NNT	Yes	Yes	Yes	Yes	NNT=1/RD	Simple proportions	No	NNT was calculated using RD between unadjusted incidence rates. Adjusted incidence rates from the Kaplan-Meier curves should have been used. For example, at one year of follow-up, NNT for the composite endpoint would be 92 from Kaplan-Meier curves, rather than 60 person-years from unadjusted incidence rates. The authors interpreted person-years as number of persons treated over one year, which is not exactly the same.
Wijeysundera et al.	Safety	Binary	Yes	Relative Risk	Primary outcome	NNTH	Patient-based NNT	Yes	Yes	Yes	No	NNT=1/RD	Simple proportions	Yes	NNT calculated as 1/RD (between groups for rates of death at 30 days). The outcome was analysed as a binary outcome, and not a time-to-event outcome as it often happens for survival outcomes. A non-parsimonious multivariable logistic regression model was developed.

Reference	Primary outcome type	Type of variable	Relative effect measure	Type of relative effect measure	NNT calculated for outcome	NNTB and/or NNTH	Type of NNT	CER presented	Time horizon defined for NNT	Confidence intervals for NNT	Methodology used to compute NNT defined in methods section	Method used to compute NNT	Source of data used to compute NNT	Adequate method to compute NNT	Comments
Nested Case-Control Study															
Etminan et al.	Safety	Time to event	Yes	Rate ratio	Primary and secondary outcomes	NNTH	Patient-time-based NNT	No	Yes	No	Yes	Relative effect measure	Relative effect measure	Yes	Logistic regression model used to compute Rate Ratios. Patient-time based NNT can be calculated this way, but the authors should have stated that the resulting NNT equals 2500 patient-years rather than 2500 patients.
Bell et al.	Safety	Binary	Yes	Odds ratio	Primary outcome	NNTH	Patient-based NNT	Yes	Yes	Yes	Yes	Relative effect measure	Relative effect measure	Yes	Not clearly defined, but we assumed that the CER of 0.3% used to compute NNTH is associated with a follow-up time of 14 days.

^{13}C-UBT: ^{13}C urea breath test; ARR, absolute risk reduction; CER, control event rate; CV, cardiovascular; NA, not applicable; OR, odds ratio; NNT, number needed to treat; NNTB, number needed to treat to benefit; NNTH, number needed to treat to harm; RD, risk difference; RCT, randomized controlled trial; RR, relative risk.

V.7.7. SUPPLEMENTAL DATA V.8 – REFERENCES FROM STUDIES INCLUDED IN THE ANALYSIS

BANGALORE, S. ET AL., 2014. BETA-BLOCKERS FOR MYOCARDIAL INFARCTION: A META-ANALYSIS OF RANDOMIZED TRIALS. AM J MED, 127(10), PP. 939-953.

BELL, C. ET AL., 2009. ASSOCIATION BETWEEN TAMSULOSIN AND SERIOUS OPHTHALMIC ADVERSE EVENTS IN OLDER MEN FOLLOWING CATARACT SURGERY. JAMA, 301(19), PP. 1991-1996.

BONGARTZ, T. ET AL., 2006. ANTI-TNF ANTIBODY THERAPY IN RHEUMATOID ARTHRITIS AND THE RISK OF SERIOUS INFECTIONS AND MALIGNANCIES: SYSTEMATIC REVIEW AND META-ANALYSIS OF RARE HARMFUL EFFECTS IN RANDOMIZED CONTROLLED TRIALS. JAMA, 295(19), PP. 2275-2285.

BRIDGE, J. ET AL., 2007. CLINICAL RESPONSE AND RISK FOR REPORTED SUICIDAL IDEATION AND SUICIDE ATTEMPTS IN PEDIATRIC ANTIDEPRESSANT TREATMENT: A META-ANALYSIS OF RANDOMIZED CONTROLLED TRIALS. JAMA, 297(15), PP. 1683-1696.

BRINKS, A. ET AL., 2011. CORTICOSTEROID INJECTIONS FOR GREATER TROCHANTERIC PAIN SYNDROME: A RANDOMIZED CONTROLLED TRIAL IN PRIMARY CARE. ANN FAM MED, 9(3), PP.226-234.

CHATTERJEE, S. ET AL., 2014. THROMBOLYSIS FOR PULMONARY EMBOLISM AND RISK OF ALL-CAUSE MORTALITY, MAJOR BLEEDING, AND INTRACRANIAL HEMORRHAGE: A META-ANALYSIS. JAMA 311(23), PP. 2414-21.

CHRISTENSEN, R. ET AL., 2007. EFFICACY AND SAFETY OF THE WEIGHT-LOSS DRUG RIMONABANT: A META-ANALYSIS OF RANDOMISED TRIALS. LANCET, 370(9600), PP. 1706-1713.

COKER, T. ET AL., 2010. DIAGNOSIS, MICROBIAL EPIDEMIOLOGY, AND ANTIBIOTIC TREATMENT OF ACUTE OTITIS MEDIA IN CHILDREN: A SYSTEMATIC REVIEW. JAMA, 304(19), PP. 2161-2169.

DENTALI, F. ET AL., 2007. META-ANALYSIS: ANTICOAGULANT PROPHYLAXIS TO PREVENT SYMPTOMATIC VENOUS THROMBOEMBOLISM IN HOSPITALIZED MEDICAL PATIENTS. ANN INTERN MED, 146(4), PP. 278-288.

ENDEN, T. ET AL., 2012. LONG-TERM OUTCOME AFTER ADDITIONAL CATHETER-DIRECTED THROMBOLYSIS VERSUS STANDARD TREATMENT FOR ACUTE ILIOFEMORAL DEEP VEIN THROMBOSIS (THE CAVENT STUDY): A RANDOMISED CONTROLLED TRIAL. LANCET, 379(9810), PP. 31-38.

ETMINAN, M. ET AL., 2012. ORAL FLUOROQUINOLONES AND THE RISK OF RETINAL DETACHMENT. JAMA, 307(13), PP. 1414-1419.

FEXER, J. ET AL., 2014. THE EFFECTS OF THEOPHYLLINE ON HOSPITAL ADMISSIONS AND EXACERBATIONS IN COPD PATIENTS: AUDIT DATA FROM THE BAVARIAN DISEASE MANAGEMENT PROGRAM. DTSCH ARZTEBL INT, 111(17), PP. 293-300.

FRANKLIN, M. ET AL., 2011. COGNITIVE BEHAVIOR THERAPY AUGMENTATION OF PHARMACOTHERAPY IN PEDIATRIC OBSESSIVE-COMPULSIVE DISORDER: THE PEDIATRIC OCD TREATMENT STUDY II (POTS II) RANDOMIZED CONTROLLED TRIAL. JAMA, 306(11), PP. 1224-1232.

GRAHAM, D. ET AL., 2010. RISK OF ACUTE MYOCARDIAL INFARCTION, STROKE, HEART FAILURE, AND DEATH IN ELDERLY MEDICARE PATIENTS TREATED WITH ROSIGLITAZONE OR PIOGLITAZONE. JAMA, 304(4), PP. 411-418.

GREEN, H., PAUL, M., VIDAL, L. & LEIBOVICI, L., 2007. PROPHYLAXIS OF PNEUMOCYSTIS PNEUMONIA IN IMMUNOCOMPROMISED NON-HIV-INFECTED PATIENTS: SYSTEMATIC REVIEW AND META-ANALYSIS OF RANDOMIZED CONTROLLED TRIALS. MAYO CLIN PROC, 82(9), PP. 1052-1059.

HALONEN, J. ET AL., 2007. CORTICOSTEROIDS FOR THE PREVENTION OF ATRIAL FIBRILLATION AFTER CARDIAC SURGERY: A RANDOMIZED CONTROLLED TRIAL. JAMA, 297(14), 1562-1567.

HEMPEL, ET AL., 2012. PROBIOTICS FOR THE PREVENTION AND TREATMENT OF ANTIBIOTIC-ASSOCIATED DIARRHEA: A SYSTEMATIC REVIEW AND META-ANALYSIS. JAMA, 307(18), PP. 1959-1969.

HOLLINGSWORTH, J. ET AL., 2006. MEDICAL THERAPY TO FACILITATE URINARY STONE PASSAGE: A META-ANALYSIS. LANCET, 368(9542), PP. 1171-1179.

IMAZIO, M. ET AL., 2011. COLCHICINE FOR RECURRENT PERICARDITIS (CORP): A RANDOMIZED TRIAL. ANN INTERN MED, 155(7), PP. 409-414.

IMAZIO, M. ET AL., 2014. COLCHICINE FOR PREVENTION OF POSTPERICARDIOTOMY SYNDROME AND POSTOPERATIVE ATRIAL FIBRILLATION: THE COPPS-2 RANDOMIZED CLINICAL TRIAL. JAMA, 312(10), PP. 1016-1023.

IMAZIO, M. ET AL., 2014. EFFICACY AND SAFETY OF COLCHICINE FOR TREATMENT OF MULTIPLE RECURRENCES OF PERICARDITIS (CORP-2): A MULTICENTRE, DOUBLE-BLIND, PLACEBO-CONTROLLED, RANDOMISED TRIAL. LANCET, 383(9936), PP. 2232-2237.

JONAS, D. ET AL., 2014. PHARMACOTHERAPY FOR ADULTS WITH ALCOHOL USE DISORDERS IN OUTPATIENT SETTINGS: A SYSTEMATIC REVIEW AND META-ANALYSIS. JAMA, 311(18), PP. 1889-1900.

JØRGENSEN, M. ET AL., 2015. BETA-BLOCKER-ASSOCIATED RISKS IN PATIENTS WITH UNCOMPLICATED HYPERTENSION UNDERGOING NONCARDIAC SURGERY. JAMA INTERN MED, 175(12), PP. 1923-31.

KAYENTAO, K. ET AL., 2013. INTERMITTENT PREVENTIVE THERAPY FOR MALARIA DURING PREGNANCY USING 2 VS 3 OR MORE DOSES OF SULFADOXINE-PYRIMETHAMINE AND RISK OF LOW BIRTH WEIGHT IN AFRICA: SYSTEMATIC REVIEW AND META-ANALYSIS. JAMA, 309(6), PP. 594-604.

KENYON, S. ET AL., 2008. CHILDHOOD OUTCOMES AFTER PRESCRIPTION OF ANTIBIOTICS TO PREGNANT WOMEN WITH SPONTANEOUS PRETERM LABOUR: 7-YEAR FOLLOW-UP OF THE ORACLE II TRIAL. LANCET, 372(9646), PP. 1319-1327.

LAZZERINI, M. ET AL., 2013. EFFECT OF THALIDOMIDE ON CLINICAL REMISSION IN CHILDREN AND ADOLESCENTS WITH REFRACTORY CROHN DISEASE: A RANDOMIZED CLINICAL TRIAL. JAMA, 310(20), PP. 2164-2173.

LENZE, E. ET AL., 2015. EFFICACY, SAFETY, AND TOLERABILITY OF AUGMENTATION PHARMACOTHERAPY WITH ARIPIPRAZOLE FOR TREATMENT-RESISTANT DEPRESSION IN LATE LIFE: A RANDOMISED, DOUBLE-BLIND,PLACEBO-CONTROLLED TRIAL. LANCET, 386(10011), PP. 2404-2412.

LEONTIADIS, G., SHARMA, V. & HOWDEN, C., 2007. PROTON PUMP INHIBITOR THERAPY FOR PEPTIC ULCER BLEEDING: COCHRANE COLLABORATION META-ANALYSIS OF RANDOMIZED CONTROLLED TRIALS. MAYO CLIN PROC, 82(3), PP. 286-296.

LEUCHT, S. ET AL., 2012. ANTIPSYCHOTIC DRUGS VERSUS PLACEBO FOR RELAPSE PREVENTION IN SCHIZOPHRENIA: A SYSTEMATIC REVIEW AND META-ANALYSIS. LANCET, 379(9831), PP. 2063-2071.

LEUNG, A. ET AL., 2011. RISK OF THIAZIDE-INDUCED HYPONATREMIA IN PATIENTS WITH HYPERTENSION. AM J MED, 124(11), PP. 1064-1072.

LIOU, J. ET AL., 2013. SEQUENTIAL VERSUS TRIPLE THERAPY FOR THE FIRST-LINE TREATMENT OF HELICOBACTER PYLORI: A MULTICENTRE, OPEN-LABEL, RANDOMISED TRIAL. LANCET, 381(9862), PP. 205-213.

LONDON, M. ET AL., 2013. ASSOCIATION OF PERIOPERATIVE BETA-BLOCKADE WITH MORTALITY AND CARDIOVASCULAR MORBIDITY FOLLOWING MAJOR NONCARDIAC SURGERY. JAMA, 309(16), PP. 1704-1713.

MAHER, A. ET AL., 2011. EFFICACY AND COMPARATIVE EFFECTIVENESS OF ATYPICAL ANTIPSYCHOTIC MEDICATIONS FOR OFF-LABEL USES IN ADULTS: A SYSTEMATIC REVIEW AND META-ANALYSIS. JAMA, 306(12), PP. 1359-1369.

MASON, B. ET AL., 2014. GABAPENTIN TREATMENT FOR ALCOHOL DEPENDENCE: A RANDOMIZED CLINICAL TRIAL. JAMA INTERN MED, 174(1), PP. 70-77.

MEROPOL, S.; LOCALIO, A. & METLAY, J., 2013. RISKS AND BENEFITS ASSOCIATED WITH ANTIBIOTIC USE FOR ACUTE RESPIRATORY INFECTIONS: A COHORT STUDY. ANN FAM MED, 11(2), PP. 165-172.

PALMERINI, T. ET AL., 2015. MORTALITY IN PATIENTS TREATED WITH EXTENDED DURATION DUAL ANTIPLATELET THERAPY AFTER DRUG-ELUTING STENT IMPLANTATION: A PAIRWISE AND BAYESIAN NETWORK META-ANALYSIS OF RANDOMISED TRIALS. LANCET, 385(9985), PP. 2371-2382.

PAREKH, T. ET AL., 2014. HYPOGLYCEMIA AFTER ANTIMICROBIAL DRUG PRESCRIPTION FOR OLDER PATIENTS USING SULFONYLUREAS. JAMA INTERN MED, 174(10), PP. 1605-1612.

PREISS, D., ET AL. 2011. RISK OF INCIDENT DIABETES WITH INTENSIVE-DOSE COMPARED WITH MODERATE-DOSE STATIN THERAPY: A META-ANALYSIS. JAMA, 305(24), PP. 2556-2564.

ROVERS, M. ET AL., 2006. ANTIBIOTICS FOR ACUTE OTITIS MEDIA: A META-ANALYSIS WITH INDIVIDUAL PATIENT DATA. LANCET, 368(9545), PP. 1429-1435.

RYAN, N.; BIRRING, S. & GIBSON, P., 2012. GABAPENTIN FOR REFRACTORY CHRONIC COUGH: A RANDOMISED, DOUBLE-BLIND, PLACEBO-CONTROLLED TRIAL. LANCET, 380(9853), PP. 1583-1589.

SHAH, E. ET AL., 2012. EVALUATION OF HARM IN THE PHARMACOTHERAPY OF IRRITABLE BOWEL SYNDROME. AM J MED, 125(4), PP. 381-93.

SHAMLIYAN, T. ET AL., 2011. SYSTEMATIC REVIEW OF THE LITERATURE ON COMPARATIVE EFFECTIVENESS OF ANTIVIRAL TREATMENTS FOR CHRONIC HEPATITIS B INFECTION. J GEN INTERN MED, 26(3), PP. 326-339.

SHEPHERD, J. ET AL., 2008. INTENSIVE LIPID LOWERING WITH ATORVASTATIN IN PATIENTS WITH CORONARY ARTERY DISEASE, DIABETES, AND CHRONIC KIDNEY DISEASE. MAYO CLIN PROC, 83(8), PP. 870-879.

SMITH, D. ET AL., 2015. COMPARATIVE EFFECTIVENESS OF STATIN THERAPY IN CHRONIC KIDNEY DISEASE AND ACUTE MYOCARDIAL INFARCTION: A RETROSPECTIVE COHORT STUDY. AM J MED, 128(11), P. 1252.E1-1252.E11.

SPIEGEL, B. ET AL., 2006. COMPARING RATES OF DYSPEPSIA WITH COXIBS VS NSAID+PPI: A META-ANALYSIS. AM J MED, 119(5), P. 448.E27-36.

SPIELMANS, G. ET AL., 2013. ADJUNCTIVE ATYPICAL ANTIPSYCHOTIC TREATMENT FOR MAJOR DEPRESSIVE DISORDER: A META-ANALYSIS OF DEPRESSION, QUALITY OF LIFE, AND SAFETY OUTCOMES. PLOS MED, 10(3), P. E1001403.

SRINIVASAN, M. ET AL., 2012. ZINC ADJUNCT THERAPY REDUCES CASE FATALITY IN SEVERE CHILDHOOD PNEUMONIA: A RANDOMIZED DOUBLE BLIND PLACEBO-CONTROLLED TRIAL. BMC MED, 10, P.14.

TESTA, L. ET AL., 2008. REPEAT THROMBOLYSIS OR CONSERVATIVE THERAPY VS. RESCUE PERCUTANEOUS CORONARY INTERVENTION FOR FAILED THROMBOLYSIS: SYSTEMATIC REVIEW AND META-ANALYSIS. QJM, 101(5), PP. 387-395.

UNGER, H. ET AL., 2015. SULPHADOXINE-PYRIMETHAMINE PLUS AZITHROMYCIN FOR THE PREVENTION OF LOW BIRTHWEIGHT IN PAPUA NEW GUINEA: A RANDOMISED CONTROLLED TRIAL. BMC MED, 13, P.9.

WIJEYSUNDERA, D. ET AL., 2008. EPIDURAL ANAESTHESIA AND SURVIVAL AFTER INTERMEDIATE-TO-HIGH RISK NON-CARDIAC SURGERY: A POPULATION-BASED COHORT STUDY. LANCET, 372(9638), PP. 562-569.

ZINMAN, B. ET AL., 2010. LOW-DOSE COMBINATION THERAPY WITH ROSIGLITAZONE AND METFORMIN TO PREVENT TYPE 2 DIABETES MELLITUS (CANOE TRIAL): A DOUBLE-BLIND RANDOMISED CONTROLLED STUDY. LANCET, 376(9735), PP. 103-111.

CHAPTER VI – GENERAL DISCUSSION

VI. GENERAL DISCUSSION

VI.1. DISCUSSION

The decisions made by regulatory authorities are of utmost importance, given that they have the responsibility to ensure that only medicines with favourable benefit-risk profiles are available for use by the society (EMA 2016a). As such, they must ensure that only safe medicines are approved, but at the same time have enough flexibility to allow into market medicines that can bring potential benefits to public health (Eichler et al. 2008), (Eichler et al. 2013).

In post-marketing, when safety concerns arise for a given drug, the overall benefit to risk balance is reassessed. There are no simple or static rules that can be followed by regulatory authorities to support withdrawal or suspension decisions. Such decisions depend on the context (Evans & Leufkens 2014). However, there are some questions that may help focusing decision-making.

When there is information coming from spontaneous reports in particular, the withdrawal of a drug should be considered once i) there is evidence of causality; ii) the adverse reaction is sufficiently serious (significant morbidity or mortality) in the context of the treated disease; iii) the risk cannot be mitigated; iv) the magnitude of harm is likely to be higher than clinical benefits (both measured in absolute terms); v) there are alternative treatments with no association with that particular harm, but with similar efficacy; and vi) the withdrawal can be managed without harming patients, health professionals and health systems (Evans & Leufkens 2014). When there is evidence from comparative studies, the causality assessment, the magnitude of the association (relative and absolute risks), and case fatality rates can be more easily calculated (Evans & Leufkens 2014). However, most studies report only relative risks, and rarely absolute risks (Nuovo, Melnikow & Chang 2002), (Alonso-Coello et al. 2016). The problem with relative risks is that the same value may correspond to a negligible or an important impact on public health. For example, a RR of 2 for a given adverse reaction may imply that one person or 100 persons out of 1000 have the event in case of a background incidence rate of 0.1% or 10%, respectively (Ma et al. 2016). Noteworthy, benefit-risk assessment must be assessed on absolute scales. Yet, the dependence of absolute risks on time frame and background incidence rates of events must also be considered (Evans & Leufkens 2014).

The assessment of benefit-risk balances is still essentially a subjective evaluation of benefits and risks of medicines, which relies mostly in clinical experts' opinions (EMA 2007), (FDA 2013a). Therefore, the conclusions and decisions made upon the assessment of the same evidence may be different among assessors depending on individual values and subjective perspectives (Walker et al. 2015). A flagrant example of divergent decisions by two major regulatory authorities is provided by the case of rosiglitazone (Mendes, Alves & Batel-Marques 2015). This antidiabetic was withdrawn from the EU market by the EMA because of its association with serious cardiovascular adverse events, but the FDA decided that rosiglitazone should continue being marketed in the USA (EMA 2010b), (FDA 2011).

In this context, regulatory authorities, pharmaceutical companies, academics and other stakeholders in the field of drug regulation initiated studies to investigate, develop and test methodologies for benefit-risk assessment (EMA 2009a), (FDA 2013a), (Mt-Isa et al. 2014), (Leong, Salek & Walker 2015), (Pignatti et al. 2015). The main aim is to increase transparency and reproducibility of decision-making process. These methodologies tend to move benefit-risk assessment towards quantitative or semi-quantitative direction, without excluding or replacing the value of clinical judgment from the process (Yuan, Levitan & Berlin 2011).

As previously noted, the benefit-risk assessment comprehends five main stages: i) planning; ii) evidence gathering and data preparation; iii) analysis; iv) exploration; and v) conclusion and dissemination. A brief description about each stage was provided in the general introduction of this thesis. The analysis stage is about the assessment of benefits and risks of medicines, including their weighing and integration, in order to provide a quantitative measure of the benefit-risk balance (PROTECT 2011), (Hughes et al. 2016). There are three main types of methodologies that can be useful during the analysis stage, namely metric indices, quantitative frameworks, or utility survey techniques. This project was aimed at studying the usefulness of metric indices, namely NNT (NNTB, NNTH, and the ratio between NNTH/NNTB, i.e. LHH), for benefit-risk assessment of medicines. Noteworthy, by the time this project began, these metrics have been recommended for further testing in benefit-risk assessment of medicines (Mt-Isa et al. 2014).

The first study of this project was aimed to study the usefulness of NNTH for post-marketing safety assessments using the case of rosiglitazone (Mendes, Alves & Batel-Marques 2015). This case was used because of two main reasons. First, because of the controversy and intense debate that was generated around the cardiovascular safety of rosiglitazone in the scientific community. Second and above all, because of the divergent regulatory actions

that were made by different regulatory authorities despite the assessment of the same clinical evidence.

This study comprised an analysis of the evidence reviewed by both the EMA and FDA to support their decisions about rosiglitazone. Data was also collected for pioglitazone, the other thiazolidinedione on the market. Further literature searches were conducted to identify additional data. The outcomes of interest included all cause death, cardiovascular death, myocardial infarction and congestive heart failure. Several comparisons were carried out for both rosiglitazone and pioglitazone based on different subgroups of studies, i.e. according to the control group (placebo or active therapy) and to the regimen (monotherapy or add-on therapy). A direct comparison between rosiglitazone and pioglitazone was also performed. Those comparisons were performed using random-effects meta-analyses and I^2 to assess heterogeneity between included studies. NNTH (95% CI) was calculated for each comparison on each cardiovascular event using pooled OR and the background annual incidence rate estimated in the Look AHEAD Research Group Trial (Look AHEAD Research Group 2013).

The overall results of the study suggested that rosiglitazone is associated with an increased risk of cardiovascular adverse events as compared to controls, including pioglitazone, across several subgroups of analysis and sources of information, i.e. experimental and observational data. Low and statistically significant values of NNTH were consistently associated with rosiglitazone versus controls for all-cause death, myocardial infarction, stroke, and congestive heart failure. The results suggested that neither rosiglitazone nor pioglitazone increase the risk of cardiovascular death. With the exception of congestive heart failure in the PROACTIVE Trial, pioglitazone was not associated with an increased risk of any other cardiovascular outcome in any set of analysis. Indeed, pioglitazone was associated with a statistically significant protective effect with regard to all-cause death in observational studies. Moreover, when directly compared to pioglitazone, rosiglitazone was associated with statistically significant lower values of NNTH for all cardiovascular adverse outcomes excepting cardiovascular death for which there was no difference between treatments.

The findings of this study, i.e. NNTH values, indicated in a consistent way across different sources of evidence that rosiglitazone presents a less favourable cardiovascular safety profile compared with pioglitazone. Although the purpose of this study is not to argue against or in favour of regulatory decisions, present conclusions are in line with those reached by the EMA. However, there are other issues which are not accounted by this

quantitative methodology, but that are considered in benefit-risk assessments. Those may include design limitations of studies used to generate risk estimates, the existence of alternative therapeutics, the relative value attributed by assessors to each source of evidence (for example, experimental versus observational studies), or lack of comparability between the characteristics of patients included in studies with those using the drug under evaluation in real-world. A brief recall about the evaluation process carried out for rosiglitazone, and a synthetic description about the views of the assessors from the EMA and the FDA may be useful to understand the influence of those issues in post-marketing benefit-risk assessment of drugs.

Uncertainties around the cardiovascular safety of rosiglitazone were noted by the time the marketing authorization was granted. The EMA requested the manufacturer to conduct a long-term cardiovascular morbidity/mortality study (later named the RECORD Trial). An increased risk of myocardial infarction and all-cause mortality was found in a meta-analysis of RCTs published in 2007 (Nissen & Wolski 2007). Two years later the results from the RECORD Trial became available, but they were not conclusive with regard to the cardiovascular risk of rosiglitazone, i.e. neither confirmed nor excluded the risk (Home et al. 2009). Further several limitations were pointed out to the design of the RECORD Trial (Blind et al. 2011), (Bourg & Phillips 2012). A new meta-analysis, published in 2010, presented results that were in line with the previous meta-analysis (Nissen & Wolski 2010). In addition, data from a large cohort study indicated an increased risk of cardiovascular adverse events with rosiglitazone versus pioglitazone (Graham et al. 2010).

Based on the available evidence, the FDA decided that rosiglitazone should continue on the market, although under a Risk Evaluation and Mitigation Strategy (REMS) programme to assure that the benefits exceeded risks in patients receiving the drug under that system (Woodcock, Sharfstein & Hamburg 2010). The fact that only one therapeutic alternative from the same class (i.e. pioglitazone) was available on the market weighed in this decision. Indeed, a possible risk of bladder cancer with pioglitazone was being investigated at that time (FDA 2010b). The assessors from the FDA considered that "when there are just two drugs in the class, and many outstanding uncertainties, maintaining some flexibility may have value for patient care" (Woodcock, Sharfstein & Hamburg 2010). Furthermore, several limitations were pointed out to the design of the RECORD Trial and an independent readjudication of end points at the patient level was requested by the FDA to clarify the findings of that study (Tucker 2013). In 2013, a FDA panel advised easing restrictions on rosiglitazone after analysing the results of the readjudication of RECORD (FDA 2013c). Nevertheless, the

decision was not unanimous, with assessors expressing that safety concerns were still present, or that a clear benefit over pioglitazone could not be identified (Tucker 2013).

The EMA decided to suspend the marketing authorization of rosiglitazone in 2010. According to the EMA's conclusions, the meta-analysis performed by Nissen and Wolski in 2010 and the observational study conducted by Graham and colleagues have particularly weighed in the final decision (EMA 2010b). Despite the uncertainties regarding the cardiovascular risk, there was no reliable evidence to refute such safety concern (Blind et al. 2011). Furthermore, the results of a retrospective cohort study indicated that 8% of the patients were being prescribed rosiglitazone despite contraindications (EMA 2010d). The EMA was unable to identify a well-defined subgroup of patients more suitable for rosiglitazone than pioglitazone (Blind et al. 2011). The EMA came to the conclusion that the benefits of rosiglitazone no longer outweighed its risks.

Interestingly, the EMA assessors might have valued more the data from observational studies than the experts from the FDA. Indeed, odds ratios and hazard ratios under 2.0, even if statistically significant, are usually viewed with scepticism and caution to support regulatory decisions within the FDA (FDA 2013b). Another issue that might have contributed to the divergent regulatory decisions might have been the difficulty in identifying a subgroup of patients where the benefit-risk ratio of rosiglitazone was positive (Pouwels & van Grootheest 2012).

One potential limitation of the present study is that only safety outcomes were assessed, assuming that there are no significant differences between rosiglitazone and pioglitazone in terms of efficacy (Mendes, Alves & Batel-Marques 2015). Nevertheless, this is a reasonable assumption given that there are few RCTs and meta-analyses suggesting that the efficacy of rosiglitazone and pioglitazone are comparable in terms of benefit on the reduction of HbA1c values and glycaemic control (Khan, St Peter & Xue 2002), (Derosa et al. 2004), (Derosa et al. 2006), (Norris, Carson & Roberts 2007), (Chapell, Gould & Alexander 2009). Further, the EMA was unable to identify a subgroup of patients that could benefit more from rosiglitazone compared to pioglitazone (Blind et al. 2011).

The inclusion of objective and validated metric indices, namely NNTH, in post-marketing drug's benefit–risk assessments could be of increased value and help regulatory authorities to make consistent decisions on drug safety. Their application may contribute to improve the interpretation of results. However, there are issues weighing in benefit-risk assessments that are not possible to express by means of these quantitative metrics.

A second study was then carried out to investigate whether there was agreement between NNTH values and withdrawals of medicines from the EU market because of safety reasons (Mendes, Alves & Batel-Marques 2016a). The hypothesis of investigation was that NNTH values for those medicines would be lower in post-marketing compared with pre-marketing. The underlying assumption was that the benefits (NNTB) would have remained constant over time for withdrawn medicines, i.e. their benefit-risk ratios became negative only because of an increase in risks for adverse events during post-marketing compares with pre-marketing.

This study analysed a 15 year period, and included medicines withdrawn from the market based on safety evidence from controlled studies since they allow to estimate event rates, relative risks and consequently the calculation of NNTH. The study comprised two periods of time: (1) pre-marketing and (2) post-marketing. Pre-marketing comprehended data obtained from RCTs used to support marketing authorizations. Post-marketing included data obtained from studies conducted after the marketing authorization that supported the withdrawal of medicines from the market. The website of the EMA was searched to identify pre-marketing documents, including EPARs, as well as documents prepared by the agency following post-marketing benefit-risk reassessments. Those reference documents were used to identify studies that supported each regulatory decision. Since data came could from more than one study for each medicine, random-effect meta-analyses were carried out to pool estimates of odds ratios with 95% confidence intervals. Those estimates were used together with annual control event rates obtained from clinical literature. If there was only one study available for the adverse event of interest, the risk estimate provided in that study was used to calculate NNTH values.

From 27 medicines withdrawn from the market, eight were included in the study for quantitative analyses: almitrine, benfluorex, nicotinic acid/laropiprant, rimonabant, rofecoxib, rosiglitazone, sibutramine, and ximelagatran.

Pre-marketing data could not be identified for few medicines on some adverse events: almitrine and peripheral neuropathy; benfluorex and heart valve disease; nicotinic acid/laropiprant and serious bleeding or serious infection; sibutramine and cardiovascular events; and ximelagatran and drug induced liver injury. In these cases, pre-marketing NNTH was considered to be infinite, i.e. an infinite number of patients would need to be exposed to a given medicine in order to encounter one additional adverse outcome of interest over a given period of time. Following this assumption, the overall conclusion is that the NNTH values decreased for all medicines from pre-marketing to post-marketing, which is in line with the regulatory decisions and therefore supporting the hypothesis of investigation.

The only exception was found for nicotinic acid/laropiprant on the risk of new-onset diabetes, since the NNTH increased from pre-marketing (NNTH=113) to post-marketing (NNTH=390). Nevertheless, the pre-marketing NNTH was non-statistically significant (NNTH=113; 95% CI: NNTH 23, NNTB 353), while the post-marketing NNTH reached statistical significance (NNTH=390; 95% CI: NNTH 245, NNTH 778). Furthermore, nicotinic acid/laropiprant was withdrawn from the market not only due to an increased risk of new-onset diabetes, but also because of myopathy, serious bleeding and serious infection. The NNTH decreased for all those adverse outcomes during post-marketing of nicotinic acid/laropiprant.

The limited access to pre-marketing data was a major difficulty in the present study. This task was easily carried out for medicines with a pre-marketing EPAR, i.e. approved by the EMA. However, in other cases, such as ximelagatran, the pre-marketing data submitted to European regulators had to be requested from manufacturers. The extent to which all the supplied data was assessed by regulatory agencies could not be assessed. In addition regulatory authorities may have reviewed other studies than those included in public assessment reports. This is a potential limitation of the study.

Further, inconsistency in the definition of outcomes across included studies may result in detection and/or selection bias and consequently affect comparisons between pre-marketing and post-marketing NNTHs. For example, heart valve disease was defined as emergent regurgitation in pre-marketing studies, and comprised all cases of hospitalization due to cardiac valvular insufficiency for any cause, mitral insufficiency, and aortic insufficiency in post-marketing studies (Weill et al. 2010), (Derumeaux et al. 2012).

Moreover, the study focused on analysing only specific adverse outcomes. Whether deterioration of efficacy profiles contributed to unbalance benefit-risk profiles of withdrawn medicines towards negative was not assessed. This issue may also have weighed on some decisions, such as in the case of rimonabant. The EMA concluded that patients tended to stop treatment earlier than they should, and that fact could lead to less benefits than expected in clinical studies (EMA 2008b).

Interestingly another case of discordance between regulatory agencies was identified in this study. Despite ximelagatran was approved in EU member states through a mutual recognition procedure, the drug was never approved in the USA due to concerns of hepatotoxicity (Astrazeneca 2003), (Jeffrey 2004). The pre-marketing studies submitted to European regulators reported no cases of severe liver injury. However, such cases were reported in other pre-marketing studies (SPORTIF III and SPORTIF V) presented to the FDA

(FDA 2004). Further, experts from the FDA relied on past experience with other drugs before deciding not to approve ximelagatran. Compared to warfarin, there were a greater number of patients on ximelagatran with a bilirubin increase in close temporal relationship to an elevated aminotransferase levels during the clinical development of ximelagatran (FDA 2004). Instances of transaminase elevation accompanied by elevated bilirubin have often predicted post-marketing serious liver injuries, including fatalities and patients requiring transplantation (Graham et al. 2001). Indeed, two patients died following serious liver failure induced by ximelagatran despite intense liver enzyme monitoring have been conducted in pivotal clinical trials. In addition, a previous experience using a risk management program based on liver enzyme monitoring had failed in the case of troglitazone (FDA 2004). Therefore, experts from the FDA voted against the approval of ximelagatran (Jeffrey 2004). The drug was withdrawn from the European market when the manufacturer was made aware of one patient who developed severe liver injury after exposure to ximelagatran has been completed within the post-marketing EXTEND Trial (AstraZeneca 2006), (Agnelli et al. 2009). Regular liver enzyme monitoring was not enough to mitigate the risk of hepatotoxicity (Agnelli et al. 2009).

The case of ximelagatran raises an important question about the usefulness of the NNTH to assess adverse outcomes that are unpredictable in nature. A similar problem applies to sibutramine, for which there were only cases of change in blood pressure and heart rate during pre-marketing studies. Thus, clinical judgment, reasoning, and scientific experience and expertise is possibly unreplaceable in such cases, i.e. when decisions have to be made solely based upon the existence of evidence about risk factors for clinical outcomes. Further, for very rare adverse outcomes the usefulness of NNTH is potentially precluded because statistically significance would be hardly achieved. A single case of drug induced liver injury in the post-marketing EXTEND Trial resulted in the withdrawal from market of ximelagatran.

Finally, despite the search within the EMA website allowed to identify 27 medicines that were withdrawn from the market due to safety reasons, only eight could be used in the study to test quantitative assessment by means of NNTH. The main reason for excluding medicines from the analysis was because safety signals were originated from other sources than controlled studies, namely spontaneous reports. Noteworthy, it has been estimated that 20% of drug safety alerts generated by regulatory authorities are exclusively based on evidence from post-marketing spontaneous reports (Alves, Macedo & Batel-Marques 2013). The use of NNTH is eventually precluded in such cases, which means that the applicability of this metric is limited in a considerable proportion of benefit-risk assessments. Nevertheless,

according to the overall results of this study, NNTH may be useful as supportive tool in benefit-risk re-assessments of marketed drugs and increase value in assisting regulatory authorities to make consistent decisions on drug safety, particularly when safety signals arise from controlled studies.

In a third study, NNT related metrics were tested to quantitatively assess benefits, risks and benefit-risk ratios of medicines approved to treat RRMS. In recent years, the therapeutic arsenal has grown substantially and clinicians face more challenging decisions when selecting treatments for their patients. Therefore, the study aimed to provide potential useful information on the clinical use of those medicines, through the estimation of NNTB for benefits, NNTH for harms and LHH as a measure of benefit to risk ratios (Mendes, Alves & Batel-Marques 2016b).

This study comprehended a systematic review of literature, according to PRISMA statement, to identify all phase III RCTs with ≥2 year duration of patient's follow-up that assessed efficacy and safety of monotherapy with approved first-line and second-line (or highly-active) medicines for RRMS. Several outcomes were used to assess efficacy namely annualised relapse rates, absence of relapses, and absence of disease progression. With regard to safety outcomes, serious adverse events, and adverse events leading to discontinuation of treatment were analysed for every medicine. Further, a pool of adverse events of interest was established for each medicine depending on its particular safety profile. Random-effects meta-analyses were performed to pool evidence from studies and estimate RR with 95% CI for outcomes of interest for every medicine. Those estimates were applied to control event rates to calculate NNTB and NNTH with confidence intervals for beneficial and harmful outcomes, respectively. Further, LHH values (=NNTH/NNTB) were calculated to determine benefit-risk ratios.

In an ideal scenario NNTB values would be as close as possible to one, and NNTH would be as high as infinite, i.e. all patients treated with a particular medicine would benefit from treatment and an infinite number of patients would need to be treated in order to one have an adverse outcome. Consequently, higher values of LHH mean better benefit-risk ratios. Benefit-risk ratios are positive when LHH is higher than one, meaning that the number of patients needed to treat to benefit from therapy is lower than the number of patients needed to treat to be harmed by therapy.

All first-line medicines (DMF, GA, ß-interferons, and teriflunomide) were compared with placebo. The lowest values of NNTB were found with IFN-ß-1a-SC for all outcomes of

efficacy, namely annualized relapse rate (NNTB 3; 95% CI: 2-4), proportion of patients free of relapse (NNTB 4; 95% CI: 3-7), and proportion of patients remaining free of confirmed disability progression sustained for 3 months (NNTB 4; 95% CI: 3-7). With regard to safety outcomes, serious adverse events were not significantly increased with any medicine versus placebo. However, adverse events leading to treatment discontinuation were more frequent with IFN-ß-1a-SC (NNTH 27; 95% CI: 5–57,495) and IFN-ß-1b (NNTH 14; 95% CI: 2-426) than with placebo. Statistically significant NNTHs were found for all medicines on several other adverse events of interest, mainly non-serious. According to LHH results, the first-line DMT with less favourable benefit–risk ratios appears to be GA. For a cut-off LHH ≥2 (i.e., the number of patients benefiting from treatment is at least twice the number being harmed), IFN-ß-1a-SC have the most favourable benefit–risk ratios.

Similar analyses were carried out for second-line therapies (alemtuzumab, fingolimod, and natalizumab), which are usually used in highly active RRMS. Studies with alemtuzumab were controlled by IFN-ß-1a-SC, while studies with fingolimod, and natalizumab were controlled by placebo. This issue demands caution when interpreting results from comparisons. The lowest NNTB values were found for natalizumab on all outcomes of efficacy. The extent to which the comparison with IFN-ß-1a-SC, rather than with placebo, prejudiced the results with alemtuzumab was not assessed. For example, alemtuzumab was considered the most effective second line medicine in reducing recurrence of relapses versus placebo in a network meta-analysis (Tramacere et al. 2015). Regarding safety profiles, statistically significant NNTHs were not found for serious adverse events with any medicine, but a protective effect was found for adverse events leading to treatment discontinuation with alemtuzumab versus IFN-ß-1a-SC. However, the lowest NNTHs were found for alemtuzumab, particularly infusion associated reactions (NNTH 6; 95% CI 1-40) and rash (NNTH 6; 95% CI 3-16). Statistically significant NNTHs were also found for the other medicines, namely abnormal liver function (NNTH 22; 95% CI 8-107) and hypertension (NNTH 32; 95% CI 13-184) with fingolimod, and lymphocytosis (NNTH 20; 95% CI 7-109) with natalizumab. According to LHH results, the most favourable benefit-risk ratios were reported for natalizumab. The less favourable benefit-risk ratios were more frequently associated with alemtuzumab, with LHH<1 when several adverse outcomes were considered against the benefit of being free of relapses.

The main conclusion is that the most favourable benefit-risk ratios were found for IFN-ß-1a-SC in first-line and natalizumab in second-line treatments for RRMS. Further, this

study allowed to infer few other important conclusions about the usefulness and potential limitations of NNTB, NNTH and LHH for benefit-risk assessment.

These metrics may be valuable for benefit-risk assessments and interpretation of results, as they reflect baseline risks of events. Those baseline risks should always be present together with the estimation of NNTB or NNTH to assure a correct analysis of results. Relying only on results expressed by means of relative measures of effect may be shortcoming. Despite the relative risk reduction on annual relapse rate versus placebo has been estimated at 44-53% with DMF (Fox et al. 2012), (Gold et al. 2012), and 33% with IFN-ß-1a-SC ([No authors listed] 1998), the number of patients needed to treat to avoid one relapse over 2 years with DMF (NNTB 7) is >2-fold the number with IFN-ß-1a-SC (NNTB 3). This information may be useful in the context of benefit-risk assessments and therefore should not be disregarded. Thus, it should be acknowledge that a higher relative risk reduction with drug A as compared with drug B (versus a common comparator in trial A and trial B, respectively) does not necessarily mean that the number of patients needed to be treated for one patient to encounter one additional outcome of interest over a defined period of time is lower with drug A than with drug B (Mendes, Alves & Batel-Marques 2016b). This is because of the baseline risk for the outcome of interest in the population included in the studies, namely the control event rate of events. The inclusion of patients with earlier or less severe states of disease in more recent clinical trials (i.e. with lower baseline risks) and with higher responses to control interventions (for example, low control event rates with placebo) contribute to the finding of higher values of NNTB with DMF compared to IFN-ß-1a-SC.

Nevertheless, there are also few situations that may preclude the applicability or the added value of using NNTB, NNTH and LHH in the assessment of benefits, harms and particularly their ratios. As already noted in the second study, the use of data from spontaneous reports to compute these metrics is challenging. For example, PML is a serious adverse event that may be life-threatening to patients and therefore must be considered in the assessment of benefit-risk ratios of medicines indicated in RRMS. Of recall, natalizumab was withdrawn from the market due to few post-marketing spontaneous reports of PML. By using total numbers of spontaneously reported cases and estimated population exposure to a given medicine, one may calculate an approximate incidence rate of PML in patients receiving that medicine. Further, the application of that result to the incidence of PML in the general population can be used to provide a rough estimate of the NNTH for the event in patients receiving a given medicine over a decided period of time. However, these estimates

will always suffer from limitations inherently associated to spontaneous reporting systems (for example, underreporting or duplicate reporting) and those associated with the estimation of exposition to a particular medicine (for example, neither all prescribed medicines are always dispensed to patients, nor all dispensed medicines are used by patients). The application of NNT related metrics deserves further exploration when data originate from spontaneous reports.

One of the main problems with NNTB and NNTH for benefit-risk assessments is that these metrics allow to compare only a single benefit and a single risk (Hughes et al. 2016). This trouble was observed in the present study since the analyses involved many criteria, i.e. several benefits and risks. Therefore a considerable amount of comparisons had to be carried out for each medicine. This fact contributes to an increased difficulty in the interpretation of results, and consequently in the achievement of robust conclusions (Juhaeri et al. 2011). Moreover, the direct comparison between a benefit and one risk implies that they are equally important. One criticism that has been made to NNTB and NNTH is that these metrics do not allow to compare outcomes with different amounts of clinical relevance (Holden, Juhaeri & Dai 2003a), (Holden 2003b), (Nixon et al. 2016), (Hughes et al. 2016). For example, the clinical relevance of avoiding a relapse is certainly different from the importance of inducing an episode of flushing over two years of treatment with DMF. Therefore the application of LHH to assess benefit-risk ratios is probably not very informative unless benefits and risks have the same clinical importance. The usefulness of using NNTB and NNTH alone to weigh up multiple benefits and risks with different relevance seems limited (EMA 2010a), (Hughes et al. 2013).

Researchers have suggested modifications to the original concept of NNT to improve the use in drug benefit-risk assessment. One extension allows to combine and weigh multiple benefits and risks simultaneously, by incorporating utilities (i.e. numeric representations of patient's preferences for specific outcomes) through relative value adjustments. That metric is the relative value adjusted number needed to harm (RV-NNH) (Holden, Juhaeri & Dai 2003a). It is calculated in the same way as NNTH, but the denominator includes the sum of differences of proportions for all adverse events of interest with adjustment for relative values. RV is the value of avoiding an adverse event relative to avoiding the target disease, i.e. RV = (1 – utility of adverse event) / (1 – utility of target disease). The RV-NNH metric was originally conceived to address the importance of considering multiple adverse events relative to a single benefit. As such, RV-NNH can be used as threshold and compared with classical NNT (or NNTB), with a favourable comparison when NNT < RV-NNH. If the approach used for harms is also applied to weigh multiple benefits, the relative value adjusted

number needed to treat (RV-NNT) can be obtained (Nixon et al. 2016). However, the use of utility-adjusted variants of NNT has been not recommended because the meaning of reciprocals utilities is different from reciprocals of proportions (Mt-Isa et al. 2014).

Nevertheless, a similar approach was tested using a case study of benefit-risk assessment (Nixon et al. 2016). The weighted net clinical benefit (wNCB) is a quantitative framework that expresses the overall difference between the sum of all weighed benefits and the sum of all weighed risks (Sutton et al. 2005), (Nixon et al. 2016). Benefits (positive contribution) outweigh risks (negative contribution) when wNCB is greater than 0. The wNCB is a particular case of multi criteria decision analysis (MCDA) that has been shown to be equivalent to the RV-NNT principle. This approach was considered simple to apply and understand for drug benefit-risk assessment, although being limited to binary outcomes and assuming linear partial value function on outcomes (Nixon et al. 2016).

The fourth study of this project was dedicated to investigate whether the calculation of NNT (NNTB and NNTH) in clinical literature was in line with basic methodological recommendations (Mendes, Alves & Batel-Marques 2017). This study is of utmost importance given that these metrics can be used to support informed decision making in clinical practice (Straus et al. 2011), (Citrome & Ketter 2013). Hence clinicians need to rely on the methodological appropriateness of the calculations of the NNTB and NNTH values reported in clinical studies, namely those published in high-impact factor medical journals, i.e. those that are more likely to influence the perceptions of clinicians about benefits and harms of medicines.

The added value of this study is two-fold. In first place, the study provides an overview about methodological aspects that should be considered for the calculation of NNT in studies with several research designs (i.e. meta-analysis, RCT, cohort and case-control study) and assessing few type of outcome variables (i.e. binary and time-to-event). The study also allowed to characterize the use of this metric in clinical research, as well as to carry out an appraisal of methods used to produce NNT in selected studies, including further discussion about limitations and implications of inappropriate use.

Secondly, the results of this study validate the methodologies that were used to calculate NNT related metrics in the other three studies included in this thesis. Noteworthy the previous studies used meta-analysis techniques to pool relative measures of effect. The results were then applied to control event rates obtained from clinical literature to derive NNTB and NNTH values for beneficial and harmful outcomes, as applicable (Mendes, Alves

& Batel-Marques 2015), (Mendes, Alves & Batel-Marques 2016a), (Mendes, Alves & Batel-Marques 2016b). This methodology was used because it was necessary to pool data from several sources in all studies. This strategy is in line with methodological recommendations (Furukawa, Guyatt & Griffith 2002), (Cates 2002), (Altman & Deeks 2002), (Deeks, Higgins & Altman 2011).

In this study, top 25 journals with high impact factor in the category of general and internal medicine were screened to identify controlled studies reporting NNT estimates to measure effects of pharmacological interventions. Data were collected to describe general characteristics of selected studies, and particularly the methods used to calculate NNT in those studies. The adequacy of calculating methods was assessed by means of comparison with basic recommendations published in scientific literature. Three references were used as main sources of such recommendations (Bender 2005), (Austin & Laupacis 2011), (Deeks, Higgins & Altman 2011). Further, a limited literature search was carried out to identify additional information.

The study included 51 publications using NNT to express pharmacological treatment effects. The research design more frequently encountered was meta-analysis (n=23; 45.1%), followed by RCTs (n=17; 33.3%), cohort (n=9; 17.6%) and case-control (n=2; 3.9%) studies. Binary variables (n=41; 80.4%) were more commonly used than time-to-event variables (n=10; 19.6%) to assess primary outcomes of selected studies. Point-estimate NNT values were sometimes presented alone, i.e. without specification of control event rate (n=14; 27.5%), time horizon (n=13; 25.5%), or confidence intervals (n=19; 37.3%).

The NNT was not calculated in accordance to basic methodological recommendations in 15 studies (29.4%). The majority of those studies were meta-analyses (n=13; 86.7%). One meta-analysis used simple proportions (i.e. raw totals from each study included in the meta-analysis were added together to derive proportions and risk differences), while 12 pooled absolute risk differences. The preferred method in meta-analysis is to produce a pooled relative effect measure to express treatment differences and derive NNT estimates. Further, one RCT and one cohort study with time-to-event outcomes also applied inadequate methods, namely because of using simple proportions rather than adjusted cumulative incidences.

The assessment of calculating methods in selected studies was performed based upon the analysis of only a limited list of basic recommendations. However, there are other issues that may influence the calculation of NNT. For example, the choice of population for risk averaging to estimate adjusted NNT from logistic regression depends on the research question of the study under evaluation (Bender et al. 2007). Other important issues include

the methods used for confidence intervals estimation or the dealing with competing risks. The assessment of these and other issues was beyond the scope of this research. Therefore, despite basic methodological recommendations have been followed in some instances, these applications may have been nevertheless inadequate because of reasons that could not be assessed with reasonable effort.

Overall, the results of this study suggested that the NNT concept is still sometimes miscalculated and misinterpreted in clinical literature. This is particularly noted in the case of meta-analyses, with more than half (56.5%) of such studies having not followed basic methodological recommendations in the calculation of NNT. This fact may lead to biased results and misleading conclusions about the effects of clinical interventions. In addition, point-estimate NNT values were presented alone in a considerable proportion of studies, without definition of baseline risks, time horizons and/or confidence intervals. The interpretation of a NNT without its context may be ambiguous and less informative for decision-making.

From the clinicians point of view these results may raise concern, since the NNT can be used to support clinical decision making processes, including the prescription of medicines. The use of calculating methods that are not appropriate to estimate NNT results may lead to biased findings and misleading conclusions (Bender & Blettner 2002), (Suissa et al. 2012), (Cates 2002), (Hutton 2000). This in turn may result in distorted perceptions about the benefits and harms of medicines, and consequently in less informed clinical decisions, which may hinder the optimal patient care. Therefore, clinicians need to rely on the methodological appropriateness of such calculations.

VI.2. REFERENCES

[NO AUTHORS LISTED], 1998. RANDOMISED DOUBLE-BLIND PLACEBO-CONTROLLED STUDY OF INTERFERON BETA-1A IN RELAPSING/REMITTING MULTIPLE SCLEROSIS. PRISMS (PREVENTION OF RELAPSES AND DISABILITY BY INTERFERON BETA-1A SUBCUTANEOUSLY IN MULTIPLE SCLEROSIS) STUDY GROUP. LANCER, 352(9139), PP. 1498-1504.

AGNELLI, G. ET AL., 2009. SAFETY ASSESSMENT OF NEW ANTITHROMBOTIC AGENTS: LESSONS FROM THE EXTEND STUDY ON XIMELAGATRAN. THROMB RES, 123(3), PP. 488-497.

ALONSO-COELLO, P. ET AL., 2016. SYSTEMATIC REVIEWS EXPERIENCE MAJOR LIMITATIONS IN REPORTING ABSOLUTE EFFECTS. J CLIN EPIDEMIOL, VOLUME 72, PP. 16-26.

ALTMAN, D. & DEEKS, J., 2002. META-ANALYSIS, SIMPSON'S PARADOX, AND THE NUMBER NEEDED TO TREAT. BMC MED RES METHODOL, 2(3).

ALVES, C., MACEDO, A. & BATEL-MARQUES, F., 2013. SOURCES OF INFORMATION USED BY REGULATORY AGENCIES ON THE GENERATION OF DRUG SAFETY ALERTS. EUR J CLIN PHARMACOL, 69(12), PP. 2083-2094.

ASTRAZENECA, 2003. EXANTA (XIMELAGATRAN) RECEIVES FIRST APPROVAL [PRESS RELEASE: 23-12-2003], LONDON, UK: ASTRAZENECA.

ASTRAZENECA, 2006. LETTER: WITHDRAWAL OF XIMELAGATRAN 36 MG FILM-COATED TABLETS: EMEA/H/C/000702 [14-02-2006], LONDON, UK: ASTRAZENECA.

AUSTIN, P. & LAUPACIS, A., 2011. A TUTORIAL ON METHODS TO ESTIMATING CLINICALLY AND POLICY-MEANINGFUL MEASURES OF TREATMENT EFFECTS IN PROSPECTIVE OBSERVATIONAL STUDIES: A REVIEW. INT J BIOSTAT, 7(1), P. 6.

BENDER, R., 2005. NUMBER NEEDED TO TREAT (NNT). IN: ENCYCLOPEDIA OF BIOSTATISTICS. CHICHESTER: WILEY, PP. 3752-3761.

BENDER, R., KUSS, O., HILDEBRANDT, M. & GEHRMANN, U., 2007. ESTIMATING ADJUSTED NNT MEASURES IN LOGISTIC REGRESSION ANALYSIS. STAT MED, 26(30), PP. 5586-5595.

BLIND, E., DUNDER, K., DE GRAEFF, P. & ABADIE, E., 2011. ROSIGLITAZONE: A EUROPEAN REGULATORY PERSPECTIVE. DIABETOLOGIA, 54(2), PP. 213-218.

BOURG, C. & PHILLIPS, B., 2012. ROSIGLITAZONE, MYOCARDIAL ISCHEMIC RISK, AND RECENT REGULATORY ACTIONS. ANN PHARMACOTHER, 46(2), PP. 282-289.

CATES, C., 2002. SIMPSON'S PARADOX AND CALCULATION OF NUMBER NEEDED TO TREAT FROM META-ANALYSIS. BMC MED RES METHODOL, 2(1).

CHAPELL, R., GOULD, A. & ALEXANDER, C., 2009. BASELINE DIFFERENCES IN A1C EXPLAIN APPARENT DIFFERENCES IN EFFICACY OF SITAGLIPTIN, ROSIGLITAZONE AND PIOGLITAZONE. DIABETES OBES METAB, 11(11), PP. 1009-1016.

DEEKS, J., HIGGINS, J. & ALTMAN, D., 2011. CHAPTER 9: ANALYSING DATA AND UNDERTAKING META-ANALYSES. IN: T. C. COLLABORATION, ED. COCHRANE HANDBOOK FOR SYSTEMATIC REVIEWS OF INTERVENTIONS VERSION 5.1.0.

DEROSA, G. ET AL., 2006. EFFECTS OF 1 YEAR OF TREATMENT WITH PIOGLITAZONE OR ROSIGLITAZONE ADDED TO GLIMEPIRIDE ON LIPOPROTEIN (A) AND HOMOCYSTEINE CONCENTRATIONS IN PATIENTS WITH TYPE 2 DIABETES MELLITUS AND METABOLIC SYNDROME: A MULTICENTER, RANDOMIZED, DOUBLE-BLIND, CONTROLLED. CLIN THER, 28(5), PP. 679-688.

DEROSA, G. ET AL., 2004. METABOLIC EFFECTS OF PIOGLITAZONE AND ROSIGLITAZONE IN PATIENTS WITH DIABETES AND METABOLIC SYNDROME TREATED WITH GLIMEPIRIDE: A TWELVE-MONTH, MULTICENTER, DOUBLE-BLIND, RANDOMIZED, CONTROLLED, PARALLEL-GROUP TRIAL. CLIN THER, 26(5), PP. 744-754.

DERUMEAUX, G. ET AL., 2012. ECHOCARDIOGRAPHIC EVIDENCE FOR VALVULAR TOXICITY OF BENFLUOREX: A DOUBLE-BLIND RANDOMISED TRIAL IN PATIENTS WITH TYPE 2 DIABETES MELLITUS. PLOS ONE, 7(6), P. E38273.

EICHLER, H. ET AL., 2013. THE RISKS OF RISK AVERSION IN DRUG REGULATION. NAT REV DRUG DISCOV, 12(12), PP. 907-916.

EICHLER, H. ET AL., 2008. BALANCING EARLY MARKET ACCESS TO NEW DRUGS WITH THE NEED FOR BENEFIT/RISK DATA: A MOUNTING DILEMMA. NAT REV DRUG DISCOV, 7(10), PP. 818-826.

EMA, 2007. COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE (CHMP). REFLECTION PAPER ON BENEFIT-RISK ASSESSMENT METHODS IN THE CONTEXT OF THE EVALUATION OF MARKETING AUTHORISATION APPLICATIONS OF MEDICINAL PRODUCTS FOR HUMAN USE (EMEA/CHMP/15404/2007), LONDON, UK: EUROPEAN MEDICINES AGENCY.

EMA, 2008. Questions and answers on the recommendation to suspend the marketing authorisation of Acomplia (rimonabant), London, UK: European Medicines Agency.

EMA, 2009. Benefit-Risk Methodology Project. Development and testing of tools and processes for balancing multiple benefits and risks as an aid to informed regulatory decisions about medicinal products, London, UK: European Medicines Agency.

EMA, 2010-a. European Medicines Agency recommends suspension of Avandia, Avandamet and Avaglim - anti-diabetes medication to be taken off the market [Press release: 23-09-2010], London, UK: European Medicines Agency.

EMA, 2010. Benefit-risk methodology project: Work package 2 report: Applicability of current tools and processes for regulatory benefit-risk assessment. [Online] Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Report/2010/10/WC500097750.pdf [Accessed 15 Feb 2017].

EMA, 2010. Cardiac profile of patients using rosiglitazone-containing anti-diabetes medicines: a study using the THIN database. [Online] Available at: http://www.encepp.eu/encepp/openAttachment/studyResult/1683;jsessionid=xoc96rfibN_WaOwx3d5Jg4PnOl-6b3pRcHtF2RFB-SM6j_9pkXRo!1497764701 [Accessed 05 Apr 2017].

EMA, 2016. European Medicines Agency - About Us - What we do. [Online] Available at: http://www.ema.europa.eu/ema/index.jsp?curl=pages/about_us/general/general_content_000091.jsp&mid=WC0b01ac0580028a42 [Accessed 27 Dec 2016].

Evans, S. & Leufkens, H., 2014. Regulatory decision-making: are we getting it right? Pharmacoepidemiol Drug Saf, 23(10), pp. 1012-1016.

FDA, 2004. Division of Cardiovascular and Renal Drug. NDA 21-686. Ximelagatran (H376/95), Silver Spring, Maryland, USA: Food and Drug Administration.

FDA, 2010. FDA drug safety communication: ongoing safety review of Actos (pioglitazone) and potential increased risk of bladder cancer after two years

EXPOSURE. [ONLINE] AVAILABLE AT: HTTP://WWW.FDA.GOV/DRUGS/DRUGSAFETY/UCM226214.HTM [ACCESSED 05 APR 2017].

FDA, 2011. FDA DRUG SAFETY COMMUNICATION: UPDATED RISK EVALUATION AND MITIGATION STRATEGY (REMS) TO RESTRICT ACCESS TO ROSIGLITAZONE CONTAINING MEDICINES INCLUDING AVANDIA, AVANDAMET, AND AVANDARYL [PRESS RELEASE: 18-05-2011], SILVER SPRING, MARYLAND, USA: FOOD AND DRUG ADMINISTRATION.

FDA, 2013-B. READJUDICATION OF THE ROSIGLITAZONE EVALUATED FOR CV OUTCOMES AND REGULATION OF GLYCEMIA IN DIABETES TRIAL (RECORD). JOINT MEETING OF THE ENDOCRINOLOGIC AND METABOLIC DRUGS ADVISORY COMMITTEE AND THE DRUG SAFETY AND RISK MANAGEMENT ADVISORY COMMITTEE, SILVER SPRING, MARYLAND, USA: FOOD AND DRUG ADMINISTRATION.

FDA, 2013-C. FDA DRUG SAFETY COMMUNICATION: FDA REQUIRES REMOVAL OF SOME PRESCRIBING AND DISPENSING RESTRICTIONS FOR ROSIGLITAZONE-CONTAINING DIABETES MEDICINES [PRESS RELEASE: 25-11-2013], SILVER SPRING, MARYLAND, USA: FOOD AND DRUG ADMINISTRATION.

FDA, 2013. STRUCTURED APPROACH TO BENEFIT-RISK ASSESSMENT IN DRUG REGULATORY DECISION-MAKING. DRAFT PATIENT DRUG USER FREE ACT (PDUFA) I IMPLEMENTATION PLAN, SILVER SPRING, MARYLAND: S.N.

FOX, R. ET AL., 2012. PLACEBO-CONTROLLED PHASE 3 STUDY OF ORAL BG-12 OR GLATIRAMER IN MULTIPLE SCLEROSIS. N ENGL J MED, 367(12), PP. 1087-1097.

FURUKAWA, T., GUYATT, G. & GRIFFITH, L., 2002. CAN WE INDIVIDUALIZE THE 'NUMBER NEEDED TO TREAT'? AN EMPIRICAL STUDY OF SUMMARY EFFECT MEASURES IN META-ANALYSES. INT J EPIDEMIOL, 31(1), PP. 72-76.

GOLD, R. ET AL., 2012. PLACEBO-CONTROLLED PHASE 3 STUDY OF ORAL BG-12 FOR RELAPSING MULTIPLE SCLEROSIS. N ENGL J MED, 367(12), PP. 1098-1107.

GRAHAM, D. ET AL., 2001. LIVER ENZYME MONITORING IN PATIENTS TREATED WITH TROGLITAZONE. JAMA, 286(7), PP. 831-833.

GRAHAM, D. ET AL., 2010. RISK OF ACUTE MYOCARDIAL INFARCTION, STROKE, HEART FAILURE, AND DEATH IN ELDERLY MEDICARE PATIENTS TREATED WITH ROSIGLITAZONE OR PIOGLITAZONE. JAMA, 304(4.), PP. 411-418.

HOLDEN, W., 2003B. BENEFIT-RISK ANALYSIS : A BRIEF REVIEW AND PROPOSED QUANTITATIVE APPROACHES. DRUG SAF, 26(12), PP. 853-862.

HOLDEN, W., JUHAERI, J. & DAI, W., 2003A. BENEFIT-RISK ANALYSIS: A PROPOSAL USING QUANTITATIVE METHODS. PHARMACOEPIDEMIOL DRUG SAF, 12(7), PP. 611-616.

HOME, P. ET AL., 2009. ROSIGLITAZONE EVALUATED FOR CARDIOVASCULAR OUTCOMES IN ORAL AGENT COMBINATION THERAPY FOR TYPE 2 DIABETES (RECORD): A MULTICENTRE, RANDOMISED, OPEN-LABEL TRIAL. LANCET, 373(9681), PP. 2125-2135.

HUGHES, D. ET AL., 2013. RECOMMENDATIONS FOR THE METHODOLOGY AND VISUALISATION TECHNIQUES TO BE USED IN THE ASSESSMENT OF BENEFIT AND RISK OF MEDICINES. [ONLINE] AVAILABLE AT: HTTP://WWW.IMI-PROTECT.EU/DOCUMENTS/HUGHESETALRECOMMENDATIONSFORTHEMETHODOLOGYANDVISUALISATIONTECHNIQUESTOBEUSEDINTHEASSESSMENTO.PDF [ACCESSED 2017 APR 05].

HUGHES, D. ET AL., 2016. RECOMMENDATIONS FOR BENEFIT-RISK ASSESSMENT METHODOLOGIES AND VISUAL REPRESENTATIONS. PHARMACOEPIDEMIOL DRUG SAF, 25(3), PP. 251-262.

JEFFREY, S., 2004. FDA PANEL RECOMMENDS AGAINST APPROVAL OF ALL INDICATIONS FOR XIMELAGATRAN, S.L.: MEDSCAPE.

JUHAERI, J. ET AL., 2011. BENEFIT RISK WAVE 1 CASE STUDY REPORT RIMONABANT. [ONLINE] AVAILABLE AT: HTTP://WWW.IMI-PROTECT.EU/DOCUMENTS/JUHAERIETALBENEFITRISKWAVE1CASESTUDYREPORTRIMONABANTOCT2011.PDF [ACCESSED 05 APR 2017].

KHAN, M., ST PETER, J. & XUE, J., 2002. A PROSPECTIVE, RANDOMIZED COMPARISON OF THE METABOLIC EFFECTS OF PIOGLITAZONE OR ROSIGLITAZONE IN PATIENTS WITH TYPE 2 DIABETES WHO WERE PREVIOUSLY TREATED WITH TROGLITAZONE. DIABETES CARE, 25(4), PP. 708-711.

LEONG, J., SALEK, S. & WALKER, S., 2015. BENEFIT-RISK ASSESSMENT OF MEDICINES: THE DEVELOPMENT AND APPLICATION OF A UNIVERSAL FRAMEWORK FOR DECISION-MAKING AND EFFECTIVE COMMUNCATION. SWITZERLAND: ADIS - SPRINGER INTERNATIONAL PUBLISHING.

LOOK AHEAD RESEARCH GROUP, 2013. CARDIOVASCULAR EFFECTS OF INTENSIVE LIFESTYLE INTERVENTION IN TYPE 2 DIABETES. N ENGL J MED, 369(2), PP. 145-154.

MA, H. ET AL., 2016. CONSIDERATIONS ON ENDPOINT SELECTION, WEIGHTING DETERMINATION, AND UNCERTAINTY EVALUATION IN THE BENEFIT-RISK ASSESSMENT OF MEDICAL PRODUCT. STAT BIOPHARM RES, 8(4), PP. 417-425.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2015. NUMBER NEEDED TO HARM IN THE POST-MARKETING SAFETY EVALUATION: RESULTS FOR ROSIGLITAZONE AND PIOGLITAZONE. PHARMACOEPIDEMIOL DRUG SAF, 24(12), PP. 1259-1270.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2016A. TESTING THE USEFULNESS OF THE NUMBER NEEDED TO TREAT TO BE HARMED (NNTH) IN BENEFIT-RISK EVALUATIONS: CASE STUDY WITH MEDICINES WITHDRAWN FROM THE EUROPEAN MARKET DUE TO SAFETY REASONS. EXPERT OPIN DRUG SAF, 15(10), PP. 1301-1312.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2016B. BENEFIT-RISK OF THERAPIES FOR RELAPSING-REMITTING MULTIPLE SCLEROSIS: TESTING THE NUMBER NEEDED TO TREAT TO BENEFIT (NNTB), NUMBER NEEDED TO TREAT TO HARM (NNTH) AND THE LIKELIHOOD TO BE HELPED OR HARMED (LHH): A SYSTEMATIC REVIEW AND META-ANALYSIS. CNS DRUGS, 30(10), PP. 909-929.

MENDES, D., ALVES, C. & BATEL-MARQUES, F., 2017. NUMBER NEEDED TO TREAT (NNT) IN CLINICAL LITERATURE: AN APPRAISAL. BMC MED, DOI: 10.1186/S12916-017-0875-8.

MT-ISA, S. ET AL., 2014. BALANCING BENEFIT AND RISK OF MEDICINES: A SYSTEMATIC REVIEW AND CLASSIFICATION OF AVAILABLE METHODOLOGIES. PHARMACOEPIDEMIOL DRUG SAF, 23(7), PP. 667-678.

NISSEN, S. & WOLSKI, K., 2007. EFFECT OF ROSIGLITAZONE ON THE RISK OF MYOCARDIAL INFARCTION AND DEATH FROM CARDIOVASCULAR CAUSES. N ENGL J MED, 356(24), PP. 2457-2471.

NISSEN, S. & WOLSKI, K., 2010. ROSIGLITAZONE REVISITED: AN UPDATED META-ANALYSIS OF RISK FOR MYOCARDIAL INFARCTION AND CARDIOVASCULAR MORTALITY. ARCH INTERN MED, 170(14), PP. 1191-1201.

NIXON, R. ET AL., 2016. A CASE STUDY USING THE PROACT-URL AND BRAT FRAMEWORKS FOR STRUCTURED BENEFIT RISK ASSESSMENT. BIOM J, 58(1), PP. 8-27.

NORRIS, S., CARSON, S. & ROBERTS, C., 2007. COMPARATIVE EFFECTIVENESS OF PIOGLITAZONE AND ROSIGLITAZONE IN TYPE 2 DIABETES, PREDIABETES, AND THE METABOLIC SYNDROME: A META-ANALYSIS. CURR DIABETES REV, 3(2), PP. 127-140.

NUOVO, J., MELNIKOW, J. & CHANG, D., 2002. REPORTING NUMBER NEEDED TO TREAT AND ABSOLUTE RISK REDUCTION IN RANDOMIZED CONTROLLED TRIALS. JAMA, 287(21), PP. 2813-2814.

PIGNATTI, F. ET AL., 2015. STRUCTURED FRAMEWORKS TO INCREASE THE TRANSPARENCY OF THE ASSESSMENT OF BENEFITS AND RISKS OF MEDICINES: CURRENT STATUS AND POSSIBLE FUTURE DIRECTIONS. CLIN PHARMACOL THER, 98(5), PP. 522-533.

POUWELS, K. & VAN GROOTHEEST, K., 2012. THE ROSIGLITAZONE DECISION PROCESS AT FDA AND EMA. WHAT SHOULD WE LEARN? INT J RISK SAF MED, 24(2), PP. 73-80.

PROTECT, 2011. WP5: BENEFIT-RISK INTEGRATION AND REPRESENTATION. [ONLINE] AVAILABLE AT: HTTP://WWW.IMI-PROTECT.EU/WP5.SHTML [ACCESSED 16 FEB 2017].

SUTTON, A. ET AL., 2005. A BAYESIAN APPROACH TO EVALUATING NET CLINICAL BENEFIT ALLOWED FOR PARAMETER UNCERTAINTY. J CLIN EPIDEMIOL, 58(1), PP. 26-40.

TRAMACERE, I. ET AL., 2015. IMMUNOMODULATORS AND IMMUNOSUPPRESSANTS FOR RELAPSING-REMITTING MULTIPLE SCLEROSIS: A NETWORK META-ANALYSIS. COCHRANE DATABASE SYST REV, VOLUME 9:CD011381.

TUCKER, M., 2013. FDA PANEL ADVISES EASING RESTRICTIONS ON ROSIGLITAZONE. BMJ, VOLUME 346, P. F3769.

WALKER, S. ET AL., 2015. A UNIVERSAL FRAMEWORK FOR THE BENEFIT-RISK ASSESSMENT OF MEDICINES: IS THIS THE WAY FORWARD? THERAPEUTIC INNOVATION REGULATORY SCI, 49(1), PP. 17-25.

WEILL, A. ET AL., 2010. BENFLUOREX AND VALVULAR HEART DISEASE: A COHORT STUDY OF A MILLION PEOPLE WITH DIABETES MELLITUS. PHARMACOEPIDEMIOL DRUG SAF, 19(12), PP. 1256-1262.

WOODCOCK, J., SHARFSTEIN, J. & HAMBURG, M., 2010. REGULATORY ACTION ON ROSIGLITAZONE BY THE U.S. FOOD AND DRUG ADMINISTRATION. N ENGL J MED, 363(16), PP. 1489-1491.

YUAN, Z., LEVITAN, B. & BERLIN, J., 2011. BENEFIT-RISK ASSESSMENT: TO QUANTIFY OR NOT TO QUANTIFY, THAT IS THE QUESTION. PHARMACOEPIDEMIOL DRUG SAF, 20(6), PP. 653-656.

CHAPTER VII – FINAL CONCLUSIONS

VII. FINAL CONCLUSIONS

VII.1. CONCLUSIONS

This thesis evaluated the usefulness of the NNT as quantitative metric for post-marketing benefit-risk assessment of medicines, using several case studies and addressing both regulatory and clinical perspectives. Four studies were conducted in order to provide answers to the initial research question. The main conclusions obtained throughout these studies are the following:

- Rosiglitazone was associated with values of NNTH that were lower than those encountered with pioglitazone for several cardiovascular adverse events across different subgroups of analysis and sources of information. Such differences were more prominent in observational studies, particularly in those that directly compared rosiglitazone with pioglitazone. These results align better with the regulatory action proposed by the EMA than with the decision of the FDA. However, quantitative methodologies do not comprehend other issues that are relevant for benefit-risk assessment, such as the importance of existing only one alternative in the same therapeutic class. Further, the conclusions about the benefit-risk ratio of rosiglitazone were also influenced by the nature of evidence, namely the weight of data from observational studies. Different points of view around this issue are also not resolved by the application of a quantitative metric.

- In general, the values of NNTH decreased from pre-marketing to post-marketing for eight medicines withdrawn from the EU market due to safety reasons. These results are in favour of the regulatory decisions made by the EMA. However, since the calculation of the NNTH can only be performed using data from controlled studies, several medicines were excluded from the analysis. Therefore, the usefulness of this metric seems limited for benefit-risk assessment when there is only evidence of safety problems originated from post-marketing spontaneous reports.

- The results of the third study illustrate the potential usefulness of NNTB and NNTH to assess benefits and risks of medicines to support clinical decisions. These metrics reflect baseline risks and absolute differences between interventions, therefore

providing information that is not immediately perceived from results of relative effect size measures. They should probably be used more often at least in addition to results obtained from the application of relative effect measures. Higher relative risk reductions may not necessarily mean less patients needed to treat to benefit one patient. A problem with the application of NNTB and NNTH in benefit-risk assessment is that only a single benefit and a single risk can be compared at each time. As such, several comparisons have to be carried out, which may hinder the interpretation of results. In addition, the comparison between one benefit and one risk implies that they have the same clinical relevance. This fact may preclude the application of LHH to express benefit-risk ratios in some assessments.

- Basic methodological recommendations are not always followed when NNT is calculated in studies published in general and internal medicine journals. This is particularly evident in the case of meta-analysis. Such issue may lead to biased estimates and misleading conclusions about the effects of clinical interventions. Further, NNT values are often reported without presentation of baseline risks, time horizons and/or confidence intervals, which reduce the usefulness of this metric for clinical decision-making.

The NNT can be effectively used to quantity benefits and risks of medicines, as well as to provide additional and useful information about the magnitude of treatment effects. From a regulatory perspective, the use of the NNT may be considered only within defined structured frameworks for benefit-risk assessment, because there are several issues weighing in the assessments that are not addressed by quantitative metrics. The application of the NNT can be problematic for weighing multiple benefits and risks with different clinical relevance. Nonetheless, whenever calculable, the NNT may be used in the benefit-risk assessment of medicines, as this metric can help to strengthen regulatory decisions. In addition, the NNT is useful for supporting informed clinical decision-making, as long as it is properly calculated. In conclusion, although the NNT does not replace other evaluations in the benefit-risk assessment of marketed medicines, it provides useful information, as well as added value in well-defined assessments.