Systematic Review Methodology in Biomedical Evidence Generation

Van der Gronde, Toon

Systematic Review Methodology in Biomedical Evidence Generation/
A.G.M.P van der Gronde – Utrecht: Freudenthal Institute, Faculty of Science, Utrecht University / FI Scientific Library (formerly published as FIsme Scientific Library, CD-β Scientific Library) no. 102, 2019

Dissertation Utrecht University. With references. Met een samenvatting in het Nederlands.

ISBN: 978-90-73346-80-2

Cover design: Vormgeving Faculteit Bètawetenschappen
Cover illustration: Marlies van der Plas
Printed by: Xerox, Utrecht

Systematic Review Methodology in Biomedical Evidence Generation

Methodologie van systematisch literatuuronderzoek in biomedische bewijsvorming

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties in het openbaar
te verdedigen op

woensdag 4 december 2019 des middags te 12.45 uur

door

Antonie Gerardus Majella Petrus van der Gronde

geboren op 21 november 1988
te Utrecht

Promotor:
Prof. dr. A.H.L.M. Pieters

Although it is widely agreed that science is cumulative, people have only very recently begun to acknowledge that scientists have a responsibility to cumulate scientifically.[1]

Lorsqu'il n'est pas en notre pouvoir de discerner les plus vraies opinions, nous devons suivre les plus probables.

When it is not in our power to follow what is true, we ought to follow what is most probable.[2]

Perché non i titoli illustrono gli uomini, ma gli uomini i titoli.

It is not titles that make men illustrious, but men who make titles illustrious.[3]

---

[1] Chalmers, I., Hedges, L. V & Cooper, H. A brief history of research synthesis. Eval. Health Prof. 25, 12–37 (2002).

[2] Descartes, R. Discours de la méthode (1637). English translation: Descartes, R, Eliot, C. W. Discourse on Method, Part 3. PF Collier & Son (1909).

[3] Machiavelli, N. Discorsi sopra la prima deca di Tito Livio, Libro terzo, Capitolo 38 (1517). English translation: Machiavelli, Niccolo. Discourses on Livy, Book 3, Chapter 38. Translated by Harvey C. Mansfield and Nathan Tarcov. University of Chicago Press (1996).

Contents

1 General introduction

In routine healthcare practice, healthcare providers (HCPs) usually have several choices regarding how to treat a patient. To decide which treatment is most likely to help a patient, tools such as systematic reviews (SRs) and clinical guidelines are available. They help to identify and critique the best available evidence for patient care.

Historically, decisions on what to provide a patient have been based on a combination of experience, anecdotal evidence, expert opinion and pharmacological reasoning.[1–3] Slowly, there has been a shift towards evidence-based medicine (EBM), which utilizes the best available evidence in combination with clinical expertise and a patient's values and expectations to guide decisions and assess which intervention is most likely to cure or prevent a disease.[4–7] The move to EBM has not only helped guide decisions that provide the best treatment for individual patients, but also improved population health and reduced costs.[8] Using effective treatments and stopping ineffective ones improves quality of life, reduces harm to patients and frees up resources for more and better healthcare services to be delivered elsewhere.

Despite significant progress of EBM, there is a gap in implementation between the science and everyday practice: many medical treatments that are not proven to be useful for patients are started every day. Even treatments that have been proven to be completely ineffective are recommended, despite the available evidence.[9]

To guide decisions in EBM, insight in the state of the current evidence is essential. However, it is difficult to keep pace with the constant development of new literature[10–15] which partially affects the gap between science and implementation. Searching the literature, selecting the relevant findings, and appraising quality requires time and skills that are not always available for the typical HCP. SRs help bridge the gap between the theoretical best treatment and the daily practice of providing healthcare to an individual patient. By using reliable techniques to find and score relevant research, they help translate the findings so that they can be used in clinical practice.

The practice of SR writing in EBM is well-established. Due to the reliable method of collecting all relevant information, meta-analyses are

possible. Next to numerical integration, these use systematic tools to appraise the quality of each study, to rule out that low-quality material skews an effect. More advanced comparisons have been built on this, including network meta-analyses, which indirectly compare treatments, and meta-regression analyses, which use trial data to find out which variables impact the treatment effects. These methods have improved evidence generation in EBM, by making findings possible that were not possible before these tools were developed.

Another development is the scoping review, which is similar to an SR, but has as its main focus to characterize a research landscape and define key concepts and definitions. This then helps with the design of a full SR, if that is required.

Health is broader than clinical medicine, though. Finding reliable advice on health-related questions is difficult for most patients and HCPs. The gap between scientific evidence and its implementation is even more prominent in biomedical disciplines that fall outside EBM, which are the primary focus of this thesis. Specifically, the areas studied are wide-ranging and include nutrition, health policy, criminology, and sports science.

For each of these disciplines, disproportionate attention is given to mainstream media, which can present false narratives or 'flashy science' that is not always evidence-based. Instead, applying the best practices of literature appraisal in medicine to these fields could have a profound impact on patients. For example, health policy, which can help or harm patients by making the right treatments available, is not consistently subjected to formal review and scrutiny. Criminology, with severe impacts on the liberty and safety of everyone, and sports science, which impacts professional athletes daily, could also benefit from more formal evaluation. Reviewing the evidence on nutrition in animal research could help speed up evidence generation and provide direction for further avenues of research.

Due to the differences between these fields, there is no one-size-fits-all approach. More research needs to be done on a tailored and standardised approach to SR methodologies for each of these disciplines which would enable better-defined treatment comparisons. Such research could also provide insights that are applicable to EBM, advancing the EBM field and related disciplines.

1.1 Overall problem definition

SRs are a type of literature review in which the author systematically appraises the evidence in papers found using a protocoled research methodology. This has become an essential tool in the synthesis of data from individual studies and is regarded as the most robust level of evidence available in biomedical research. SRs are not consistently used in health-related disciplines other than medicine, though implementing the SR methodology could vastly improve the life of patients.

In EBM, SRs have been accepted as an essential tool to find the best options, and their methodology has been somewhat standardised. This is needed, as there are many choices in the design of an SR, which can have a significant impact on the outcome. The methodology for SRs in other health-related disciplines, however, is not yet fully developed. Here, we will discuss how to tailor the methodological design of an SR to find the most reliable outcome, and what best practices are for conducting an SR in health-related sciences.

1.1.1 Goal

To contextualise the choices in the design of an SR and evaluate the appropriateness of each method for a specific health-related research question.

1.1.2 Research question

What choices and considerations can be made in the design of an SR for a health-related research question, and how do they impact the outcome and conclusion?

1.1.3 Hypothesis

Though SRs in medicine can provide lessons on how to perform an SR, each research question requires a tailored methodological approach to find reliable results. Like healthcare providers should weigh the evidence and consider how applicable the evidence is to their patient, systematic reviewers should consider how appropriate a method is to answer a specific research question.

1.2 What is EBM?

To answer the research question, it is necessary to contextualize the background of EBM and SRs. This will help put the relevance of this research question in context. The next paragraph discusses the history and current state of EBM.

1.2.1 Definition of EBM

There is some debate about the best way to define EBM. The original publication by the EBM working group in JAMA called EBM a paradigm shift, which shifts medical practice away from "*intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision making"*, focussing instead on "*efficient literature searching and the application of formal rules of evidence evaluating the clinical literature."*[1]

This leaves room for interpretation. The problem with the term EBM is that it seems to suggest that any other form of medicine is not evidence-based.[16,17] It also suggests disqualifying clinical practice before the introduction of EBM, which caused resistance from practitioners. These issues resulted in EBM being far more narrowly defined, reducing it to "*only using RCTs [randomised controlled trials] as evidence, regardless of the patient's details*", or too broadly, as "*the best way to practice medicine."*[16]

A more nuanced definition is the following:[14] "*Evidence-based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients."* So, despite how the name is often interpreted, EBM is intended to be the integration of the best available evidence with the specific situation-related data of the patient, to select the best available treatment for that individual.[5,17]

That statement, "the best evidence," suggests that not all evidence is equally robust. EBM specifically defines which types of evidence are most valuable, which is typically represented in the shape of a pyramid: the evidence pyramid.

1.2.2 Levels of evidence: the evidence pyramid

Many biases influence the perception a healthcare provider (HCP) has of which treatment is best, and independent, well-analysed evidence is needed to guide HCPs in their decision making.[18] Different studies leave room for different biases and reliability in the evidence they generate,[3] so differentiating the reliability of evidence based on the source is essential.

For example, very experienced physicians were convinced that heart failure patients with a narrow QRS complex would benefit from cardiac resynchronisation therapy (CRT), based on ample clinical

experience and medical literature, including observational studies and case reports.[18,19] The large randomised controlled trial (RCT) that they set up to prove that the intervention was effective did not provide the evidence they were looking for, though. The intervention turned out to be dangerous for this group, with a statistically significant increase in mortality in some subgroups.[20]

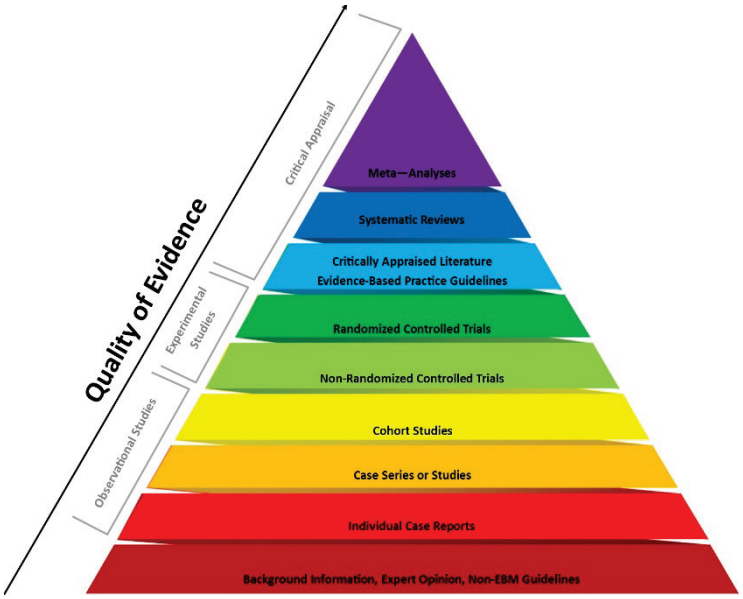

Figure 1. EBM pyramid.

The evidence pyramid is designed to help guide which study is more likely to represent the 'true' effect of an intervention. Research methods that are most valued have a low risk of bias, so are not likely to have a systematic deviation from the ground truth, leading to an overestimate or underestimate of an effect.[21] Those are placed at the top of the pyramid, whereas methods that have a high risk of bias are at the bottom.[14] There is no consistent ranking of studies based on the number of patients. A high-quality clinical trial might have just several or even one subject,[22] whereas very large studies might suffer from significant methodological flaws.

EBM focusses mainly on interventional studies such as RCTs, which generally have a lower risk of bias than observational studies.[23] The decision to use a specific drug is then not left to the prescriber, who might base their choice on patient characteristics which could influence

outcomes, but on randomisation according to stratification factors, which does not have that bias.[23] Therefore, RCTs have become the gold standard for EBM, and are regarded as the most reliable sources of evidence.[5,7,14] So a HCP should first search for a RCT with a patient population similar to the ones that are seen in practice.[24]

But even RCTs carry a risk of type 1 and type 2 errors, or of finding an effect when there is not one or vice versa, so the most reliable way to determine the effect of a treatment is to look at several studies and compare the results. The only way of reliably finding all relevant studies is using a systematic search. That is why systematic reviews are consistently at the top of the evidence-based medicine hierarchy.[4,16,25–27]

Types of studies that are lower on the hierarchy, such as large observational studies, cohort studies, and case-control studies, might be less reliable to assess the effect of a treatment. They are more likely to produce results that are due to confounding factors or biases (e.g. allocation bias, reporting bias or a healthy cohort bias).[28,29] They can give strong indications though, so have an important function in early sharing of the available knowledge and exploring future research opportunities. They also allow for research in areas where RCTs are not feasible. For example, finding risk factors and quantification or identification of adverse effects:[5,13,23,29] a RCT into the risk of lung cancer due to smoking would not be ethical. A case-control study is more appropriate for that research question.

This might make it appear as if the evidence pyramid is a well-defined and generally accepted framework. It has been, and still is, highly debated. In the next paragraph, the main criticisms are discussed.

1.2.3 Criticism on the evidence pyramid

The main criticism on the evidence pyramid is that it seems to suggest that RCTs are always better than observational research.[26] In fact, Archie Cochrane stated that to enable HCPs to make correct choices, they need insight in the benefits and costs of a treatment, which "*can really only be obtained by an adequately costed RCT.*"[30,31] There are, however, examples where observational research is the best possible method of finding the true effect of a treatment.

First, it is not always ethical to perform a RCT. For example, as a paper in the BMJ showed, there is no good evidence for the use of parachutes to prevent injury or death after jumping out of an airplane

("gravitational challenge").[28] Though this effect is widely accepted, no RCTs have been performed (one has been published though,[32] but is of dubious quality). Similarly, nobody would insist performing a RCT for an intervention that has already been widely established, such as physicians washing their hands before delivery.[33]

Second, observational studies can typically be performed faster, cheaper, and can answer a wider range of research questions than RCTs.[29,34–36] Some populations are too small to allow for a RCT,[5] and other populations simply cannot afford to wait for an RCT to research the best option.[14,37] So observational research is often a useful step to identifying an effect, which may then be replicated in a RCT. RCTs cannot reasonably be used for determining risk factors[36] or adverse events,[38] so observational studies remain needed. Also, in nature, observational studies are closer to the real-life setting than RCTs, which can make results more likely to translate into practice.

Third, in some settings RCTs are difficult to execute. For example, for surgical techniques the allocation of treatments cannot be blinded easily and in an ethical way, and crossover between arms is typically high. So observational data from comparable patient groups with follow-up is more reliable and feasible in some examples.[3]

Fourth, the fundamental criticism of observational studies is that identified and unidentified confounding factors could be responsible for the observed effect or lack thereof. A large study compared the outcomes of observational studies with RCTs, and found that the effect was similar in 2 out of 19 comparisons.[29]

Also, the evidence pyramid suggests that SRs and meta-analyses are always more reliable than RCTs. The quality of SRs is, however, heavily dependent on the quality of the evidence they identify.[39] This in term depends on methodological properties of an SR, such as the approach of searching and selecting, timing and interpretation.[4] A well-performed SR on reliable RCTs yields more reliable evidence than a single RCT, but to put SRs on the top of the pyramid seems to suggest all SRs are superior.

In conclusion, the evidence pyramid is a general guide to which types of research are most reliable at providing evidence for a clinical question. More research methods than just RCTs are required for EBM.[40] There are many examples of when the highest standard is not available,

though, and other research methodologies than RCTs have a fundamental role.

1.2.4 Why use SRs for EBM?

For healthcare practitioners, it is important to use the most reliable evidence on which treatments to use. To find this, a practitioner would have to go through the available literature and critically appraise the quality of individual publications and determine the level of evidence. With over 4800 biomedical and health journals indexed in Medline alone,[6] indexing over 10.000 articles per week, it is virtually impossible to stay up to date.[10–15] On top of that, clinical duties demand a significant amount of time and put more pressure on not spending too much time on literature.[6,30] This increases the appeal of using only the most recent or easiest obtained publications, which might reduce the quality of the used evidence.[24]

To allow HCPs to work evidence-based, SRs offer a solution. They take the pressure to search the literature, select and appraise papers, and draw a conclusion from potentially conflicting sources off of the practitioner.[30] Having a reliable summary of an entire body of literature available in an accessible format makes evidence-based medicine possible.[11,41]

1.2.5 What is an SR?

There is wide agreement on the scope and remit of systematic reviews. The Cochrane collaboration[11] describes a systematic review as a research method in which the investigator:

> *"attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made."*

It further defines main characteristics of SRs:

- a clearly stated set of objectives with pre-defined eligibility criteria for studies;
- an explicit, reproducible methodology;
- a systematic search that attempts to identify all studies that would meet the eligibility criteria;

- an assessment of the validity of the findings of the included studies, for example through the assessment of risk of bias; and
- a systematic presentation, and synthesis, of the characteristics and findings of the included studies.

The EMTREE[42] entry used in EMBASE describes SRs as "*Used for studies that systematically summarize all relevant evidence pertaining to a defined health question, and including items identified as such by the author*."

MeSH[43] defines Review [Publication Type] as "*An article or book published after examination of published material on a subject. It may be comprehensive to various degrees and the time range of material scrutinized may be broad or narrow, but the reviews most often desired are reviews of the current literature. The textual material examined may be equally broad and can encompass, in medicine specifically, clinical material as well as experimental research or case reports. State-of-the-art reviews tend to address more current matters. A review of the literature must be differentiated from HISTORICAL ARTICLE on the same subject, but a review of historical literature is also within the scope of this publication type*."

1.2.6 Other uses of SR

Conclusions from SRs are typically the basis for the development of guidelines on how to treat specific conditions in an evidence-based manner.[5,44] Granting agencies increasingly use SRs to justify the allocation of funds to a specific project.[45] Only a well-conducted SR can credibly summarize an entire subject, and thereby expose an evidence gap, or give a rationale for a specific intervention.[6,10,46] Similarly, policymakers are using SRs to identify which interventions are available and which ones are most likely to work.[3,45,47] Furthermore, SRs help in the design of new studies by giving a complete overview and can help estimate effect size in a specific population.[40] Finally, SRs can be used to provide input into cost-effectiveness studies.[44]

1.2.7 Criticism on EBM

From the start, EBM and its predecessors have known substantial opposition. The term itself, evidence-based, raised opposition, as it seemed to dismiss all other medical practice as not evidence-based and inferior.[13,16]

In the time of Pierre Louis' (1787-1872) numerical method, physicians did not agree with disregarding their own clinical experience and traditions because of someone else's numbers.[17,48] Those numbers were often derived from single hospitals, so whether their lessons could be extrapolated was debated.[17] Whether confounding factors, such as specific treatments being chosen for more severe cases, influenced the result was also uncertain. Especially when comparing hospitals numerically, advanced hospitals would be likely to deal with more severe cases, so that their mortality data was unfairly skewed.[17,49]

Furthermore, critics felt that the numerical method yielded averages and risks, but did not treat individual patients, so was not applicable to clinical practice.[17] Medicine was thought to be about individual patients, not about groups.[17]

In Louis' time, no formal hypothesis testing had been developed yet. Though the numbers might have seemed to clearly favour one treatment, there was no method of showing that a specific effect was due to chance or not.[17]

Though the Lancet favourably reviewed the Cochrane collaboration (see paragraph 1.3.12), the journal saw EBM as an internal threat to the autonomy of the physician, as it wrote in an editorial. It mostly saw overpromising and rigid views as a problem for medicine, though.[50] Many of the old arguments are still valid today and are mentioned regularly.[51] The tendency to treat patients on the basis of a population, and not on the clinical judgement of the physician, is seen as detrimental to medical practice.[16,18,51] Critics fear of reducing medical practice to 'cookbook medicine,' reducing the options for physicians after all best treatments per disease have been defined.[3,36] They feel there is no single right way to practice, however, as all patients are different.[13] EBM relies too heavily on clinical trials, they argue,[16] and does not take the complexity of the clinical reality into account.

This argument is partially based on an exaggeration of the goal of EBM, though. In the earliest publications about EBM at McMasters university, EBM was explicitly used to complement clinical judgement.[1] The authors identified the shortcomings in only trusting evidence, and stated that many patients and conditions would never qualify for thorough clinical trials.[1] In such cases, clinical reasoning and judgment

would still be required to decide on the applicability of the evidence and identify the best treatment.[5,14,51–53]

To put this debate into context, and highlight the intertwined history of EBM and SRs, the next paragraph will discuss several highlights. This is by no means an exhaustive overview, but it will provide background to better understand the current debate on EBM.

1.3 History of EBM and SRs

The start of the movement of evidence-based medicine is often attributed to a teaching program at McMaster University, Ontario.[7] Though their publication[1] did coin the term,[17] and laid out how to integrate this paradigm in a postgraduate medical educational setting, there are relevant predecessors.[14]

Similarly, SRs and meta-analyses were considered an essential tool for making EBM feasible from the start,[1,53] but their history, too, goes further back. This paragraph will give an overview of the most important developments of the intertwined history of EBM and SRs.

1.3.1 James Lind and scurvy

1.3.1.1 The trial (1740s)

James Lind (1716–1794) is often seen as the father of the clinical trial. In 1747, he worked as the ship surgeon for the British navy, when many sailors were diagnosed with scurvy. Lind divided 12 similarly ill crewmembers into groups of two, and treated them, isolated from the main population, with one of six treatments: cider, a weak acid, vinegar, seawater, nutmeg and barley water, or oranges and lemons. Thought the supply of fruits ran out on day six, the effect was already clear, as the two sailors in the fruit treatment arm were working again. The others remained 'weak in the knees'.[54,55] Though his approach was mostly informal and intuitive, it did show the relevance of testing a hypothesis for biomedical evidence generation.

1.3.1.2 The review on scurvy (1753)

Lind wanted to know if others had done similar work to his and compare their findings. This led to the first documented SR, in 1753, on the treatment of scurvy.[45,46] He briefly described how he had to sort through evidence to reach a full and impartial view.[55] He found many irrelevant documents, so "*it was necessary to remove a great deal of rubbish.*"[55] Though in this example Lind performed a trial before he searched the

literature, it does show that it was already clear then that it was useful to read research from others and integrate the findings of individual trials to create guidance on a treatment.

1.3.2 The great debate (1830s)

France in the early 1800s was the scene of a long debate on from what medicine should obtain its knowledge. There were two main traditions: experimental physiology and clinical epidemiology.[56] Where the first was focussed on the importance of experimentation and translation from animal testing, the second developed methods to look at aggregate populations to determine treatments. The next two sections give examples of the main proponents and their views.

1.3.2.1 The numerical method: Pierre Charles Alexandre Louis

During an epidemic of diphtheria, Pierre Charles Alexandre Louis (1787–1872), who worked as physician to the Tsar in Odessa, Russia, realised that he did not have the knowledge to properly address this disease. Eager to learn more, he realized that even influential physicians did not know much, and he decided to learn on his own, from the patients themselves.[17,57]

At La Charité Hospital in Paris, France, he started collecting observations from two wards as a clinical clerk.[17,57,58] He tabulated the outcomes of 2000 patients over seven years. Among these patients, there were 77 with pleuropneumonia and pneumonitis,[17] for which he recorded at which day after the onset of the disease they received the best available treatment, which at the time was bloodletting.[17] While 44% (18/41) of patients who received the treatment within the first four days died, only 25% (9/36) of those who received it later died.[17] He did not perform any formal statistical hypothesis testing, as this was not available at the time. He concluded that the data did not support the use of bloodletting to reduce duration of the disease or mortality, and advised a cautious use of the therapy.[57,59]

James Jackson (1777–1867), the writer of the preface, introduced Louis' work to the readers, and wrote that "already in his hands medicine […] begins to assume the form of an exact science."[57] His methodology was not welcomed by the field at the time though, and seen as a shift in authority from physicians to mathematicians. Despite Louis' results and ardent following, his methods were regarded with scepticism and eventually disappeared from clinical practice.[17,58,60,61] During the spring

of 1837 there was a particularly heated debate, both in the press and in person, between the proponents (mainly Louis) and opposers (mainly Xavier Bichat, Claude Bernard, and Francois Magendie) of medical statistics.[58] The most prominent commentator, Claude Bernard, would focus on Louis' results and methodology as the target of his criticism.

1.3.2.2 Claude Bernard and Experimental Medicine

As soon as Louis published his reports, Claude Bernard (1813–1878) opposed the methodology. In 1865, he published his book "Introduction à L'étude de la Médecine Expérimentale," in which he elaborated on his stance in the debate with Louis.[62]

He wanted to move away from eminence, and focus more on scientific proof for a theory, "even when the [wrong] theory is supported by great names and generally accepted."[63] He also acknowledged that there were what we now call levels of evidence, and that the distinction is important: "Theories are only hypotheses, verified by more or less numerous facts. Those verified by the most facts are the best, but even then they are never final, never to be absolutely believed."[63]

He disagreed with Louis on what those facts and truths look like, though. In contrast to his mentor, Laplace, he did not see calculus of probabilities as a useful tool.[64] He had fierce criticism on the numerical method of Louis, as he doubted the use of aggregate data for individual patients. Bernard saw each patient as an individual experiment, which could not be compared to other patients or average outcomes: if a procedure results in a mortality of 2/5, that means "absolutely nothing scientifically and provides no certain basis for a new operation, for you do not know if this new case will be among the cured or among the dead."[62,64]

Scientific reasoning to discover the cause of clinical outcomes was more important, in his opinion, than collecting data on the actual outcomes of a treatment.[63] Following this, he did not support the extrapolation of findings from other patients to guide the decision of how to treat a new patient.

Though he believed that statistics were a useful tool for the advancement of medicine, and he praised the work of PCA Louis, he did not see it as fundamental and advocated moving forward from statistics towards deterministic medical understanding ("scientific medicine" or "experimental medicine").[61,65] Aggregates and means obscured the true

biological phenomena that explain a treatment response, so should not be used.[66] Some of his comments are still heard as criticism of EBM today.

In the great debate, it seems like Louis' numerical method (which we would call clinical epidemiology now[36,58]) has prevailed over Bernard's experimental medicine. The experiments in animals and medical reasoning, as recommended by Bernard, and the symptom-based criteria for diagnoses are essential tools for drug development and disease understanding, but the clinical trial is still considered the ultimate test to decide on the treatment for patients and the guiding principle in EBM.[17]

1.3.3 Pre-world war II SR tools

In the 19th century, statistical progress yielded new tools for evidence synthesis, and several impressive reviews were published.[45] But it was really in the 20th century that research synthesis began to emerge.[45] In 1904, Karl Pearson (1857–1936) published a paper on the effect of a typhoid vaccine, integrating data from 11 studies from across India and South Africa and explaining their differences.[67] In 1932, Ronald A. Fisher (1890–1962) published a paper on a method to combine p-values from studies to obtain an overall p-value.[68,69] These tools allowed for the integration of data from individual trials.

1.3.4 A non-randomised trial: Archibald L. (Archie) Cochrane (1940s)

Another notable example is when Archie Cochrane (1909–1988) was imprisoned by the German forces in Greece during the second world war. Due to his education as a physician and his German language skills, he was appointed as the medical director of the prisoner of war camp, serving thousands of other soldiers.[2,70] To convince German army officials of the possibility to cure oedema in the legs, which many prisoners developed, he set up a trial.[30] He allocated 20 soldiers to either have their food be supplemented with a yeast extract or have the normal rations. Though it was not blinded and the numbers were small, the intervention worked and he convinced the army officials to supplement the meals with the yeast extract.[71] Later, Cochrane described the trial as his "*first, worst, and most successful trial*."[71] He continued his training as an epidemiologist, partially under tutelage of Sir Bradford Hill (1897–1991).[30,70]

After his training, he worked at the MRC Pneumoconiosis Research Unit in Cardiff. Here he tested whether the development of pneumoconiosis progressive massive fibrosis in coal miners was due to tuberculosis.[70] But his reputation is mostly based on his later monograph *"Effectiveness and Efficiency: Random Reflections on Health Services"*[31] about how only healthcare that has been proven to work should be provided in a healthcare setting with limited resources. He argued for clinical trials for all interventions.[70]

1.3.5 The first RCT: Austin Bradford Hill and the Medical Research Council (1940s)

Sir Austin Bradford Hill (1897–1991) is usually credited with the first modern RCT, for his work at the MRC.[3,23,72] He researched the impact of streptomycin for treating pulmonary tuberculosis, comparing the outcomes to similar patients given the standard of care, bed rest.[73,74]

Streptomycin was produced in the USA, where inclusions in clinical trials dropped once streptomycin became more widely available. In the UK, however, the supply of streptomycin was limited, due to the failure of local manufacturers to produce it. Demand for tuberculosis treatment was high, so there was a need to select patients and leave other patients untreated. Randomisation relieved the doctors from making this decision, and made the distribution fair.[75]

The most common method of randomisation at the time was alternation, which alternates the allocation of patients to treatment or control group based on the order of inclusion. As the physician typically knew whether a patient was receiving the treatment or control, this could influence the decision to include the next patient or could make a physician delay the inclusion of a favoured patient to allow for allocation to a preferred slot. Another method was the coin toss, using pairs of patients with similar traits. This trial used sealed envelopes, with only the name of the hospital and the patient number marked on the outside, which the central office would open after inclusion to determine if the patient would receive streptomycin and bed rest or bed rest alone.[75,76] This was a multi-centre trial, providing high-level results from several sites. In their publication, the researchers admit that they do not present an exhaustive clinical analysis, but it does allow more significant conclusions than any single centre with a smaller number of patients can reach.[74]

1.3.6 SRs in the post-world war II era

SRs slowly became more common in the post-war period.[77] In 1954, Daniels and Hill published a paper exploring the combined lessons from three clinical trials on streptomycin for pulmonary tuberculosis, including the clinical trial that was discussed in paragraph 1.3.5.[78] The integration of studies was regarded as a type of research in its own right in the 1970s.[79] Academics started to highlight the need for better integration in social sciences to make the cumulative knowledge that had already been acquired more available.

The need to help HCPs assess and digest evidence was clear even in the 1980s, when the scientific publication rate was at a fraction of what it is today. A quality check of 50 published SRs on basic methodological attributes (explicit purpose, sources, in- and exclusion criteria, discussion of limitations, quantitative integration, clear summary) concluded that major methodological gaps were common. None of the found SRs qualified for all quality criteria.[80] It was only in the 1990s that the use of SRs really increased and that the methodology was standardised.[68,81]

1.3.7 Forest plot

The forest plot, a method of visualising the effect sizes found in different studies to easily compare them and show the overall effect, was first used by Taylor, Parker, and Langenberg in 1969. They used the plot to illustrate the evidence from 12 publications on e/h, an atomic constant called the "fine structure constant."[82] This method was adopted in medical sciences. To make the larger horizontal lines, which indicate less reliable and less informative estimates, less visually dominant over the precise estimates of short lines, little boxes reflecting the inverse of the variance of the estimate were added in the 1980s.[45] This allows for the easy presentation of findings from separate studies, so that HCPs can quickly understand the field and apply the findings.

1.3.8 Meta-analysis

The term meta-analysis was introduced in a 1976 lecture by Gene Glass (1940–) on the need for better synthesis of results in research. The goal of MAs is to find the real effect of a treatment by combining multiple studies. This could be missed in an individual trial, due to chance, but the combination of several trials reinforces the strength of the results, even if the individual trials differ.[83]

The term was adopted quickly by some, but as the interpretation differed and the term gradually lost a clear meaning, opposition formed[84] (authors referred to "mega-silliness" in 1978,[85] "Statistical trickery" in 1991,[86] "Shmeta-analysis" in 1994[87] and to "statistical alchemy for the 21st century" in 1995[88]). This was both a critique on the use of the term and on the value of integrating research to reduce the effect of a single study's bias and create more reliable estimates of the effect of a treatment.[45] There were also doubts about whether diverting the intellectual investments needed to perform a high-quality meta-analysis from primary research was worth it.[52,88]

1.3.9 Funnel plots

In 1984, the first funnel plots were introduced as a tool to detect publication bias.[89] If there is publication bias, meaning that studies with positive results are more likely to be published than those with neutral or negative results, this can sometimes be identified with a funnel plot.[44] This plots a measure of precision, usually either the sample size (the number of patients) or the standard error, on the horizontal axis, and the effect estimate for each trial on the vertical axis.[10]

If large studies show a robust but mediocre effect, smaller studies should randomly show outcomes that are much more positive and negative than the large studies.[24] This would form a V-like shape, or a funnel, hence the name. If this symmetry does not exist, and there is a gap in the V, publication bias might be an issue. This is typically caused by not-published smaller studies with a low or negative effect, leading to the overestimate of the total effect.[10] A sizeable number of studies is required for this, though, otherwise the plot is inconclusive.[90] The use of funnel plots was further developed and popularised by Egger and colleagues.[91] This tool again helped to make findings of multiple studies more reliable and easier to understand.

1.3.10 Cumulative meta-analyses

In the early 1990s, cumulative meta-analyses were introduced. This method sorts qualifying trials by the date of their publication, and estimated what could have been know anytime a new trial was published given the evidence available up to that point.[92,93] This allowed for an estimate of when the superiority of one treatment over another was established beyond doubt, and showed clearly that large trials performed after that date would not have been needed if the

investigators had started with a meta-analysis of the available evidence.[94] Though replicating findings in different populations to make sure they are robust is valuable, it is debatable how ethical it is to treat large numbers of patients with an inferior treatment after the best option has been established.

Though SRs and meta-analysis slowly became more common and more standardised, there was no organisation tasked with performing them and physicians were still dependent on whether anyone had felt responsible for systematically analysing and integrating the research for a given clinical scenario. In 1979, Archie Cochrane, mentioned before in paragraph 1.3.4, wrote in his contribution to a book:[95] "It is surely a great criticism of our profession that we have not organised a critical summary, by speciality or subspecialty, adapted periodically, of all relevant randomised controlled trials."

1.3.11 Evidence-based medicine: McMaster University, Ontario (1990s)

The term "evidence-based medicine" is often attributed to the medicine department of McMaster University, in Hamilton, Ontario, Canada. There, researchers developed a new method of teaching medicine, which they called EBM.[7] The core tenet was that students should not accept the lessons of teachers as absolute truth, but search and assess the literature themselves to find answers to clinical questions, similar to how a systematic reviewer would work.

In their key publication,[1] they described that EBM *"de-emphasizes intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research."*

To say that the EBM movement was born at McMasters is overly simplistic, as many other had already pioneered the main tenets of this methodology. Their method of teaching did allow for an entire generation of physicians to appreciate the need for proper evidence appraisal, therefor this phase is an essential part in the acceptance and further development of EBM.

1.3.12 The Cochrane collaboration

A significant step in the development of SR and meta-analysis methodology is the Cochrane collaboration, named after Archie

Cochrane.[2,3] The collaboration was founded in 1993 by Ian Chalmers and 70 colleagues.[30,45,70] The goal of the Cochrane Collaboration is to create and disseminate up-to-date reviews of RCTs of healthcare interventions in order to help HCPs make informed decisions. In 1995, they launched the Cochrane Database of Systematic Reviews, a key component of the Cochrane library.[45] The Cochrane collaboration is widely recognised for providing the highest-quality reviews for medical interventions.[15]

Nowadays, the medical literature review methodology has been further refined and developed to allow for a reliable overview of the treatment options. Statistical methods to appropriately estimate treatment effects have been developed, vastly improving the usefulness of SRs.[46]

1.3.13 Forefront of EBM today

Despite the history and convincing benefit, EBM still faces opposition in specific areas of the world and niches of medicine.

1.3.13.1 Traditional medicine

In parts of the world where other cultures and beliefs play a large role in sickness and healing, the debate between traditional medicine and EBM is still ongoing. In practice, many patients choose to seek help from both types of doctors. For example, as traditional African witchdoctors state they offer more comprehensive healing and align better with the beliefs of their patients, they still have an active following.[96,97] Additionally, witchdoctors often work at the community level, so are more available in the village, whereas EBM is only practiced in larger cities, requiring substantial travel time and costs.[96]

1.3.13.2 Surgery

Another area where EBM has only sparsely penetrated the standard of care is surgery. Though there are movements to require evidence for new surgical procedures, existing ones are not tested against each other. This is particularly troubling for new developments, such as robot-assisted surgery. There was no evidence to suggest that robot-assisted surgery leads to better outcomes than standard procedures, but robots have been widely implemented. When recently a clinical trial and a cohort study suggested it actually reduces overall survival,[98,99] the FDA warned against its use, but nothing prohibits surgeons from recommending robot-assisted methods to their patients anyway.

There are difficulties in measuring the impact of surgery. For example, placebo surgery is not always ethical, but required to control for the impact of the ritual of the operation. Similarly, the training of the surgeon plays a role in the outcome of the surgery, which is hard to control for, especially for new techniques. It is not impossible to measure though, and observational studies could help provide evidence. Without the evidence to support a new procedure, is a wide implementation – let alone its costs – really justified?[100]

1.4 Need for SRs to apply new findings

This overview highlights the slow yet gradual implementation of EBM into daily practice. A major part of the challenge in this implementation was judging the evidence on quality and making the conclusions available to the field in an easily digestible format. This has been addressed by SRs, which allow practitioners to benefit from the knowledge without investing the time to assess it. As such, SRs are now widely accepted as an essential contribution to the implementation of EBM.

Other research areas could benefit from the insights gained in this area. This would allow for faster implementation of research findings, leading to improved outcomes for patients. Furthermore, experience obtained in these other areas could enrich the methodological possibilities for SRs in EBM. In the next chapter, the available choices when writing an SR are discussed, which will be applied to specific questions in the chapters after that.

2 The method of systematic review

This chapter describes the methods for writing an SR. First, we will discuss the different review types and their properties, before we dive into the details of SRs. Then we will show some examples of where this can go wrong, to highlight the relevance of these design choices.

2.1 Types of Reviews

There are several types of systematic reviews, all with their own strengths and weaknesses. The choice between the available types of reviews needs to be made on a case-by-case basis, depending on the most appropriate methodology. The next paragraphs will give details on the most common types.

2.1.1 Narrative review/expert review

Narrative reviews, also called expert reviews or traditional reviews, are qualitative, non-systematic overviews of the literature. One or more experts in a field write a paper based on their experience and insights. They have an important role in continuing education for readers and providing up-to-date insights.[12] In narrative reviews, there are usually no details on the protocol, search methodology, hypothesis,[10] selection criteria and appraisal of evidence, making the reader more dependent on the integrity of the authors.[12] Compared to systematic reviews, the topic can be much broader[44,101] and the structure less rigid, allowing the author to explore a problem in the most appropriate or clinically relevant way.[90]

This lack of methodology makes it prone to bias.[10,102] Experts might think they know a research area well, and have all the relevant experience to have a well-formed opinion on a topic, but their knowledge can be subject to selection bias, publication bias and recall bias. They are also likely to overemphasize their own research if they published on the same topic before. This can lead to cherry-picking to include citations that concur with the opinion of the author, while leaving out studies that contradict them.

As the search is not performed in a systematic way, it is not possible for readers to audit the methodology and make sure that no incorrect selection criteria, like agreement with the view of the writer, were used. Narrative reviews are often not characterised as evidence-based,[10] and considered less likely to find the unbiased truth than

systematic reviews due to the skewing effect of the opinion of the experts, who might think they know what the answer ought to be.[18,103]

One high-visibility example of the problems with narrative reviews is when Nobel prize winner Linus Pauling argued that vitamin C helps improves longevity and makes patients feel better.[103] The book, structured as a narrative review, was populated with selective citations to support this finding.[104] When a systematic review was performed on the effects of vitamin C, and all relevant studies were included, no clinical benefit was found.[105]

A more recent example is when a narrative review on treatment options for prostate cancer was published, and it claimed that olaparib, a new drug, was FDA-approved for prostate cancer.[106] No references to this statement were mentioned. The drug had at this point been tested in a phase 2 trial for prostate cancer but had not been FDA-approved for this indication. Statements like these could make HCPs assume that this drug is a good treatment choice for patients with prostate cancer, even though there is no science to back it up. This could harm patients.

Narrative reviews do have merit, though. They are useful to gain a broad overview as background information,[10] for research areas where there is too little literature to synthesise in a meaningful way, or where there is a need for an outside view. Also, narrative reviews allow for the synthesis of literature and clinical experience, and the inclusion of considerations that might not follow from published literature alone.

2.1.2 Systematic review

A more structured type of review is a systematic review (SR). SRs use a predefined and reported methodology to answer a clearly defined question, so that the methodology can be repeated independently.[6,107] More details on all the specific considerations on how to perform an SR are detailed in paragraph 2.2.

Unlike narrative reviews, SRs reduce bias by using a non-biased and explicit search methodology.[6,108] This results in a non-selective group of papers, which can be objectively assessed for their quality and then be used together to answer the question. This makes properly conducted systematic reviews evidence-based.[10]

Systematic reviews can synthesise the quantitative data if the data allow for that.[90] If there is a statistical quantitative summary, the systematic review is regarded an SR with meta-analysis.[6,10]

For emerging research areas with varying methodologies and where SRs have not yet been performed, a scoping review might be appropriate to summarize the evidence. This is a relatively new method, but gaining in popularity.[109] It aims to quickly map the key concepts and sources of evidence,[110,111] with a much broader question and allowing often diverse types of evidence.[112] The goal is more to make an overview of the available evidence than to synthesize this evidence, as an SR would do. Following that, assessment of biases and quality are not undertaken.[113] The results can then be used to design a formal SR, if appropriate, but can be of value on their own as well.[111]

2.1.3 SR with Meta-Analysis

If several trials on a specific treatment have been conducted and published with conflicting results, HCPs might not be able to decide on whether a specific treatment is better than an alternative or not. Meta-analysis is a technique to pool appropriate data from individual primary studies to come to a single estimate of the effect of a treatment,[4,6,11,107] which provides a more precise estimate of the effect of a treatment than a single trial.[44,103] This can be done on study level, using aggregate data from the selected studies, or on patient-level data, if the primary datasets are made available and analysed.[114] Such individual participant data meta-analyses offer better data quality and allow for more analyses than study-level analyses, but also require more data availability, expertise and time.[114]

Meta-analyses also allow for the estimation of consistency across studies, and can provide insight on how reliable the effect of a treatment is.[11] Systematic reviews with a meta-analysis are often referred to simply as meta-analyses, though they are a subtype of systematic reviews.[6]

Meta-analyses provide a framework for the appraisal and integration of studies with similar research questions and measures into one estimate, so it is vital that all relevant studies are found, and a proper systematic search is performed. They are appropriate for answering concrete questions about treatment options, but less suited to answer more complex questions such as how or why something works.[46]

Though the principles of meta-analysis seem straightforward, the value and reliability depends on what was done and how the practical challenges, ambiguity and nuances are dealt with.[6,24] Often, meta-analyses cannot overcome the limitations that are posed when trying to combine trials with different structure and choices into one estimate. If the original studies are biased, or too different to compare, the meta-analytic method can reduce variability due to chance and provide an estimate, but cannot correct the flaws of the original trials.[24]

Further methods to quantitatively assess treatment effects have been developed. Network meta-analyses use both direct and indirect comparisons between treatments to establish which one is most effective or harmful. This can yield more reliable estimates of treatment effects than individual trials.[107]

Another tool for synthesising data is the meta-regression analysis. This is a technique used to test whether there are different effects in different subgroups (e.g., age, gender, ethnicity, disease stage,…) of studies.[10,115] This can help explain what differences between studies have been caused by. Both methods depend heavily on the availability of reliable and comparable data, so are not always applicable.

2.1.4 Strengths of Systematic Reviews

The main strength of systematic reviews is that they provide quick answers to complicated questions with many facets to HCPs who do not have the time or expertise to assess the literature as detailed as an SR writer would.[24,103]

SRs make data from research synthesis available faster and more objectively than textbooks.[12] For example, it took ten years before evidence from several RCTs on thrombolytic therapy for acute myocardial infarction was included in textbooks.[116] Similarly, prophylactic lidocaine continued to be recommended in textbooks 20 years after all of several randomized controlled trials had suggested it did more harm than good.[116] This delay in evidence integration likely cost many patient lives, though the used evidence was widely available, stressing the relevance of quick and thorough dissemination of evidence through SRs.[45] SRs allow for a much more up-to-date synthesis of the available evidence than books do.

SRs take longer to develop than primary research, which allows for the critical appraisal of evidence. This means SRs can give a nuanced

overview of new data and can compare findings of different trials. Therefore, SRs present the optimum in timing and reliability for HCPs, which is part of why they are considered the highest quality of evidence.[41]

2.1.5 Weaknesses of Systematic Reviews

The main weakness of SRs is publication bias. SRs are dependent on published data at the time of the search, so there might be relevant data that never was published, or which was delayed and is not found. It is well-established that significant findings are more likely to be published than non-significant findings.[117–120]

Publication bias, or non-reporting bias, can derive from a lack of interest of journals to publish negative or inconclusive findings. It can also be a result of delay by the sponsor, intentionally or not. If results are not favourable or new, they might not be perceived a priority by the investigator or sponsoring company.[117,121] This means a lower capacity of time, writers and data scientists can be allocated, resulting in delayed analysis of the data, writing of the paper and publishing. Also, unfavourable results might be less likely to be published in the journal of choice, as editors might not prioritize them and peer reviewers might find them less appealing,[122] resulting in delays by selecting another journal.[121]

When companies find data that is negative or inconclusive, which they can see as not favourable to their commercial interest, they might want to delay the publication intentionally.[123] This would delay lessons learned from undergoing critical appraisal, inclusion in SRs and guidelines, and finally reaching clinics. Companies might also publish only primary data, whereas for positive findings they would publish secondary analyses and pooled analyses to highlight the findings and increase the presence of the findings in the scientific literature.[124] This is why Melander *et al.* coined the term "evidence b(i)ased medicine."[125] Systematic reviews are essential to use tools which can detect publication bias in a research field.

To try to prevent companies from intentionally delaying data that they would rather not see published, the guidelines set up by the International Committee of Medical Journal Editors (ICMJE) and the third edition of Good Publication Practice (GPP3) guidelines both specifically mention this.[126,127] ICMJE advices editors not to consider the result of a

trial in their decision to publish, to give inconclusive or negative trials a fair chance of getting published – with mixed results.[127] GPP3 advices the use of publication plans in commercial settings, to allocate resources in advance of trial read-outs to ensure the results have no impact on the timing of publication.[126] Both also advice editors to not consider, or only consider as exemption, trials that have not been registered on clinicaltrials.gov or EudraCT. Registration of observational research on ENCePP is also becoming more common. Forcing all trials to be registered at the start makes it hard for companies to withhold publication and makes them accountable.[117]

In a recent paper, investigators found that papers are still selectively delayed in phase 3 oncology research. Positive results were reported in peer-reviewed journals with a median of 272 days after the press release announcing completed analysis, whereas negative results took a median of 407 days.[121] This highlights that despite the efforts of professional organisations, delays in publication are still a relevant source of time-lag bias in systematic reviews and meta-analyses.[44]

Another paper reported on how different sponsors perform on adherence to the European Commission guideline, which aims to ensure publication of all trials within 12 months of completion. It found that half of the registered trials on EU Clinical Trials Register had not been published within 12 months, and that research with a commercial sponsor was substantially more likely to have results posted in time than research with a non-commercial sponsor.[128]

Another common weakness of SRs is the narrow research question, which makes it hard for HCPs to find an SR that answers the exact question they have. This makes it tempting to use the conclusions of an SR with a similar subject, and extrapolate the results to their patient. Next to a decent search, clinical reasoning remains essential to determine whether the results of an SR can be used for a specific patient.[44]

2.2 Steps in SRs

SRs are systematic because they use a specific set of steps to identify and integrate the right literature sources. In this paragraph, an overview of the different steps and considerations is given.

2.2.1 Define the question and outcomes

The research question should be clearly defined before the search and selection is initiated.[10] One of the dangers when writing an SR is that it is tempting to include anything that is remotely relevant, so it is important to start with a clear questions that defines the scope and has identifiable boundaries. This will help with considering whether a document that is found in the search should be included or not.

How narrow or broad a question is formulated will determine how focussed the resulting paper is, but also how generalisable the results are.[4,10] If the question is too broad, the number of results will be too large to manage.[129] If the question is too narrow, there will not be enough literature to answer it, or there might not be a meaningful integration of the results.[4]

One way of making sure that the research question for an SR on a biomedical subject is appropriate is using the PICOS criteria: [4] participants, intervention, comparison, outcome, and study design.

One cause of bias in SR writing is answering a different question than originally posed, thus using a non-complete set of literature.[10] Using a clear research question and focussing on answering that throughout the development of the paper prevents this. Questions other than the primary research outcome can be assessed and answered post hoc, but with the necessary caution.[4]

It is also helpful if the research question is interesting to the investigator, so the critical assessment of the literature is not a dreaded chore.[129] Furthermore, it should be considered if the answer would contribute to the progress of the scientific discipline.[129]

2.2.2 Control for duplicates

Once the research question is defined, it is necessary to see if anyone else has already answered this exact same question. Just like for RCTs, every investigator wants to be unique and replication of a prior study is less likely to produce novel and innovative ideas.[81] Though some duplication, maybe two or three SR with meta-analysis per indication, would be beneficial, highly duplicative work can be a waste of research resources. For example, 21 meta-analyses on the benefit of statins for prevention of atrial fibrillation after cardiac surgery seems redundant, and their conclusions did not offer new insights.[81] Similarly, 185 meta-analyses in seven years on the effect of antidepressants on depression seems

excessive. Though this might represent a marketing strategy more than a genuine search for evidence,[130] it is important to make sure that no efforts are duplicated. If no good SR has been written recently on the proposed research subject, writing one could be a valuable addition to the scientific progress.

2.2.2.1 PROSPERO

It is possible that the same research question is being answered by someone else while one is working on the SR. If that other investigator published earlier, it might be hard to publish the paper and the contribution to the progress of the field is diminished significantly. To prevent this, it is possible to search the PROSPERO database to make sure the effort is not duplicative.[90]

Furthermore, to prevent anyone else from starting the writing of an SR during the development of the manuscript, registering the protocol on PROSPERO makes it clear to other investigators that someone is already working on the subject. This will also yield a unique identification number, which journals often require before publication.[131]

The PROSPERO registry was developed by the Centre for Reviews and Dissemination at the University of York in England. Their objectives were to reduce unplanned duplication of systematic reviews, to provide transparency in the review process and to minimize reporting bias.[131] Selective publishing of SRs, similar to selective publishing of clinical trials, creates bias and leads to overestimation of the effect of a treatment in the literature.[4]

PROSPERO is only focused on reviews with at least one outcome that has direct relevance to the care of a patient, so is not relevant for all SRs.[90,132] Furthermore, only a small fraction of SRs register in advance on PROSPERO,[81] so the lack of duplicate registration does not mean that there is no other SR in development elsewhere.

2.2.3 Search the literature

2.2.3.1 Sources

Literature searches generally involve searching digital databases of literature, with the goal of finding all relevant publications.[133] The selection of databases is important to make sure as many relevant

publications as possible are found, and not too many irrelevant publications are screened.

The most used datasets are Medline through Pubmed, Embase, Scopus and Google Scholar. Additional databases with specific focusses can help improve the search results. Examples are PsychInfo, Cumulative Index of Nursing and Allied Health Literature (CINAHL), British Nursing Index, Web of Science, TRIP and Cochrane CENTRAL. Also other literature sources, such as congress abstracts, regulatory product information documents, clinical trials databases (clinicaltrials.gov, EudraCT, WHO ICTRP, industry-led study registers per company), patent records, grey literature (BIOSIS Previews) and news sources (LexisNexis) can be included for completeness.[10,103,133,134]

PubMed is a database that searches records in MEDLINE and other databases, which is maintained by the United Stated National Library of Medicine and National Institutes of Health. It includes just over 30 million references as of August 2019, and has records going back to 1946.[135] Embase is more focussed on drugs and pharmaceutical research. It is maintained by the commercial party Elsevier and includes all MEDLINE records plus non-MEDLINE indexed records, totalling 31 million references dating back to 1947.[136] Scopus, also by Elsevier, contains 69 million records as of August 2017 and holds records as old as 1788.[137]

These different databases are structured differently, though, through their use of thesauri. A thesaurus is a hierarchical reference structure that groups terms with similar meanings together. Where PubMed uses MeSH (medical subject headings) terms to index their records, Embase used Emtree headings. The structure of these thesauruses and the choices of tags to allocate to each record differ.[134] Scopus does not use a thesaurus and indexing at all.[137] Google Scholar also uses no thesaurus, but does tailor the results to previous searches and frequently updates the algorithms, making the searches less reproducible.[134] As such, the results of the same search might not yield the same records.

A problem with all systematic searches is publication bias, as discussed in paragraph 2.1.5. Studies with inconclusive or negative results might not get published as fast as other studies or might get published in journals that are not indexed in all databases. In practice, it

is rarely feasible to find all relevant studies in just one database.[4,133,138] To create a comprehensive overview, searching multiple independent datasets is essential.[139]

2.2.3.2 Search term

Defining the search term is always a balance between the sensitivity and the specificity (recall and precision): the search needs to find all the relevant sources but should not find papers that are irrelevant.[133,134] The only way to reach optimal sensitivity is searching for everything that has ever been done, whereas the best way to reach perfect selectivity is to make the search so restrictive that hardly any material is found at all. These two goals need to be balanced, by creating a search string that is tailored to the research question.

Another consideration is the time and resources required for selecting the relevant material in all the hits and dissecting the included records. Most commonly, a search string is a combination of keywords based on PICO criteria (Population, Intervention, Comparison, Outcome) with synonyms and abbreviations, connected with Boolean operators (AND, OR, NOT), truncation (e.g. depress* for both depression and depressive) and exploded search terms to keyword that are linked to the selected. For complex questions, help from a librarian with expertise in searching medical datasets can be essential.[41]

2.2.3.3 Limits

Most search engines provide the option to apply limits to the search, based on tags and metadata that have been assigned to a record by the indexers. This is an easy way to limit the number of irrelevant documents and reduce the number needed to read (NNR).[133] Appropriate limits need to be selected cautiously, though, because it could cause the accidental omission of relevant material.[140] This also highlights the need for using different search engines, as they might each allocate different tags to each record, with differences in precision and timing.[133]

2.2.3.3.1 Language

A reliable limit that can help reduce unwanted findings is the language filter. If the investigator will not include findings reported in a language that is not understandable to them, a filter on language will not unnecessarily reduce the number of hits. If the research subject is something that is more prevalent in a specific geographic region, and the inclusion of a local language is possible due to the linguistic skills of the

investigators, inclusion of this language will likely improve the number of relevant reports.[10]

2.2.3.3.2 Timeframe

Another way of making a search more specific to the research question is the timeframe. If the subject is a new drug, limiting the search to the date the drug was first reported on in public makes sense.[10] This also reduces the NNR, reducing the amount of work.

2.2.3.3.3 Publication type

Depending on the desired inputs, it is possible to limit a search based on publication types. For example, congress abstracts might give information about clinical trials months or years before they are published in scientific journals. Though this makes more up-to-date information available, including it in a search also means that the data has not been properly examined yet, that there might be errors in the results and that the methods cannot be assessed properly.[4]

An umbrella review would only include other reviews and meta-analyses, to provide a very broad overview of a research area. A search for such a review would be limited to systematic reviews, meta-analysis and sometimes narrative reviews.

Using the limits to find specific publication types, such as only randomized controlled trials, will vastly reduce the number of irrelevant hits but might come at a price. Search engines such as PubMed and Embase have a delay in their indexing, so many RCTs might not have that tag allocated at the time of their inclusion. So limiting to RCT could make a search not produce the most recent RCTs.[134] Furthermore, not all articles with that tag will actually be RCTs, as these tags are not always accurate.[81]

2.2.4 Additional sources

To make sure no articles are missed in the systematic search, additional methods can be used in addition to the systematic search in scientific databases. One additional way of finding key papers is manually searching key journals.[141] This requires knowledge about the most relevant journals, though, and might leave out key papers that were published in other journals.

Another method is reading the references used in published studies, and considering including them.[103,129,134] Scanning the references

from those studies can provide even more relevant publications.[105] This is known as reference scanning or citation scanning, and relies on the selection of previous authors.[133] As such, the danger of this method is that it might repeat the selection bias of another paper without scrutiny and assessment, so it should always be combined with a complete systematic search and be performed for multiple references.

To further reduce the risk of publication bias, additional sources can be found by reaching out to authors of presented abstracts who might be working on a full journal article.[103] The danger of this strategy is that abstracts have not undergone peer review, and might contain inaccuracies.[10]

Additionally, many systematic reviews add non-systematically found articles to their sources to provide a comprehensive overview. This can be based on advice from reviewers and co-authors, or very relevant material that is published just outside the limits of the paper. This needs to be reported appropriately, to not obscure selective additions.

2.2.5 Select the relevant material

As James Lind already stated in 1753 (see paragraph 1.3.1.2),[55] even after a carefully designed search many irrelevant documents need to be excluded. The selection heavily affects the outcome, so it is important that it is done right:[25] garbage in, garbage out.[44,84] Which papers are included and which ones are not needs to be carefully noted, including the reasons for in- or excluding them, so that this is reproducible and does not result in a bias.[24] The best method to guarantee objective selection is to have separate individuals select, and reconcile their differences by conversation.[10,41] This reduces the variability of reviewer judgement.[133] Cohen's kappa coefficient (κ) can be used to calculate the degree of agreement. If this is very high (close to 1), it indicates a good agreement on the selection. A very low κ, on the other hand, of close to 0, indicates little agreement in the selection, and can expose differences in interpretation in the research question or eligibility criteria for the selection.

It is important to make sure that duplicate patient populations are excluded from the analysis. Clinical researchers often publish interim results or secondary analyses of trials, which would skew the results to assign too much weight to a single trial if they are all included.[10] The

easiest way of identifying duplications is looking at the author list or patient enrolment dates. The paper with the longest follow-up or most applicable outcomes should be included, and the other publication discarded.[4]

2.2.5.1 Conflict of interest

The publication of a trial in a well-regarded scientific journal might seem to be a seal of approval and a guarantee for the scientific quality,[142] but it is important to assess the setup and findings carefully. There is ample evidence that industry-sponsored studies are more likely to produce positive results for their products than independently funded trials.[8] In fact, clinical trials with commercial funding have been found to be more likely to report a positive outcome for the sponsor than non-sponsored studies in orthopaedics,[143] psychiatry[8,144] or general investigative drugs.[123,145,146] This indicates that it is relevant to assess the quality of a sponsored publication properly, and consider how reliable the findings are. A study might be more reflective of the specific research agenda of a company than of the needs of a patient.[8,147] For example, pharmaceutical interventions for osteoarthritis in the knee were much more studied than surgical interventions, reflecting the commercial interest, though this might not be a fair representation of the available treatment options.[148] Another clear example of the effect of a sponsor is the analysis by Heres *et al.* on "why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine" in head-to-head trials.[124]

Though industry-sponsored trials are usually of high methodological quality,[124] there are ways to skew the outcomes in favour of the drug that the sponsor has a stake in. Common ways of skewing the results in head-to-hear trials include:[8,44,123,124,146,149]

- dose-response bias: using a lower dose in the comparator arm than for the equivalent investigative arm to improve the comparison in favour of the investigative drug;
- the reverse, using too high a dose of the comparator, so that the new drug seems less toxic;
- limiting the dose ranges of the comparator drug so that dose reduction due to side effects or titration is not possible;
- using an inferior drug, formulation or placebo when a better treatment is available;

- focussing the results and conclusion on a secondary outcome when the primary outcome is not met, which might not be powered appropriately;
- being selective in the outcomes that are reported in the results and conclusions, and selective in mentioning the p-values;
- reporting separately on individual centres where the outcome was favourable in a multi-centre trial;
- selecting subpopulations, trial length or outcome measures that are known to favour one drug, but might not be representative for the entire population, and phrase the conclusion as if it is;
- presenting select *a priori* unplanned analyses on select subpopulations of the trial, with the risk of false positives;
- report percentage change in ranking scale outcomes, instead of absolute difference or clinically meaningful outcomes;
- performing a non-inferiority trial with a low significance threshold or a population that is too small to show a difference, showing equivalence instead of superiority;
- reporting intention-to-treat or per-protocol analyses, using favourable definitions of each, whichever one is more convincing;
- not adjusting the significance level for multiple testing;
- ignoring hierarchical testing (alpha spending).

Some argue that industry-sponsored studies should not be included in SRs at all, as they are prone to a bias towards the sponsor's drug. Studies funded by public funds are then painted as more reliable, so excluding industry-sponsored studies altogether from SRs is tempting. Both publicly-funded and sponsor-funded studies have their distinct drivers and risks of biases, though.[124]

Studies with public funds need to publish an output, so will make sure they can find something that can be published. Given the increased likelihood of positive results to be published, as compared to negative results, investigators are incentivised to find something, or present the findings as more significant than they are.

The replication crisis is a clear example of that. Many findings, mainly in social sciences and psychology, turned out not to be replicated when others tried.[150] There were many reasons for this.[151,152] Sample sizes are typically small.[153] Difficulties in blinding and lack of resources for blinded review caused more favourable outcomes. Also, investigators

selectively reported positive outcomes, even when the study was not powered to find this. Some individual investigators even produced data without running the claimed experiments.

Many solutions that could have addressed this already exist. At study-level, improvements such as consistently pre-registering studies and implementing blinded data review could have prevented some of these issues. These methods have been developed and are routinely used in other research settings with registrational intent. At a meta-level, the lack of effort to replicate studies is worrying. If replication studies, even if they are negative, are rewarded as initial studies are, it would make the landscape more likely to implement this. Also, a more consistent research methodology and the more frequent use of systematic reviews to compare findings, flagging lack of replication and flagging publication bias would have helped.

Commercially-sponsored studies are more regulated than publicly funded ones and tend to focus on the possibilities for regulatory approval and reimbursement of a drug. This will make sponsors avoid any methodological shortcomings that might face scrutiny from regulators and reimbursement agencies, as this could damage the earning potential and waste their investment. They will selectively report secondary outcomes and exploratory analyses, though.

Also, publicly funded trials are usually not as fast and comprehensive as industry-sponsored funds, due to the more limited nature of public funds.[124] Finally, the pharmaceutical and medical device industry funds approximately two thirds of all major RCTs that are published in scientific journals,[8,144] leaving a substantial gap if industry-sponsored studies were to be excluded.

In conclusion, both commercially funded and publicly funded have their biases and methodological fallibilities. For now, it will remain necessary to include both industry-funded and publicly funded studies in SRs. Scrutinising the method and interpretation of the findings, and assessing a publication bias, is essential.

2.2.6 Summarize or calculate

A good review does not just give an overview of what has been published, but also discusses the findings, identifies problems and suggests research opportunities.[129] In the case of a meta-analysis, the final outcomes should be calculated with a confidence interval, notes to the

reliability and heterogeneity testing.[44] A meta-regression analysis even allows for the identification of factors that correlate with the found effect.

Heterogeneity (as opposed to homogeneity) refers to how much the trials differ from each other in terms of variation in the true effect sizes and in factors that could have influenced those effect sizes. A perfectly homogenous set of trials would be duplicates of each other. Differences in patient population or intervention between the included studies could impact the outcome of the trials, so estimating whether the trials could be compared directly is necessary. This can be done with the Q statistic (for within-study variance), the I^2 (for the ratio of variability of results among studies to total observed variation, so from 0 to 1) or the τ^2 (for between-studies variance).[10,24]

It is important to look at the biological sense of combining studies to answer one question. As an early opponent of meta-analyses points out: "6-month-old children, small dogs, large cats, and huge fish can be regarded as a homogeneous group" if only weight is taken into account.[88] He also pointed out the misconception that high heterogeneity leads to a wider applicability.[88]

Heterogeneity is inherent to the nature of EBM, where different research groups try to replicate findings and extend them to different geographies, populations, follow-up periods or compare with different drugs.[24,44] Reasons for heterogeneity can be identified and new hypotheses can be identified with these findings,[103] which could lead to identification of subpopulations or different settings.

Most SRs with meta-analysis use aggregate data from included studies, but some request the patient-level data from the original investigators and perform their analyses on a combined dataset.[24] This makes it possible to power for subpopulation that have not been reported on in the primary population, and is an important check of the original analysis. Some reanalyses of clinical trial data have found different results than the original primary publication.[81]

2.2.7 Conclude, assess bias, discuss

An essential part of reporting the results is estimating the validity of those results. As publication bias is a typical source of unreliable estimates (see paragraph 2.1.5), assessing the likelihood that this bias occurred is key.

One way of doing that is scoring individual studies on their quality. One system that does so is Grading of Recommendations Assessment, Development and Evaluation (GRADE),[154] which rates studies based on their properties for use in guideline development.[155] RCTs start as high-quality, whereas observational studies start as low-quality evidence, but this can be upgraded or downgraded based on study design and risk of biases.

Cochrane recommends the use of a 'Risk of bias' table. This scores each included study on sequence generation (selection bias), allocation sequence concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), selective outcome reporting (reporting bias) and other potential sources of bias. This allows for the estimation of an effect using only highly reliable studies, so that the overall effect is reliable.

For clinical trials, there are clinical trial registries that show that they have been planned or performed. This makes it possible to search for unpublished trials, which would have otherwise not been found.

There is no way of knowing about unpublished animal trials, however. Therefore, the data could be biased towards studies that are considered relevant for publication by investigators and editors. A visual method of showing that publication bias is likely is a funnel plot (as discussed in paragraph 1.3.9).

When the results are calculated, including the reliability of these results, this can be used to answer the original research question and formulate a take-home point. Most readers will only read the abstract, or if they open the full paper only the last paragraph of the conclusion, so formulating a clear statement improves the delivery and chances of making it to clinical practice.[4]

2.2.8 Report

To make sure that all parts of the method of an SR or meta-analysis are appropriately described, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and checklist were published.[4,90,132] This superseded the methods that were used before then, the Quality of Reporting of Meta-analyses (QUORUM)[4] for meta-analyses of randomised trials and the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group for meta-analyses of

observational studies.[10] The QUORUM statement has improved the quality of reporting, as one study showed.[156]

Nowadays, most journals ask for a PRISMA flowchart,[157] which provides details of the selection and screening procedure, and a PRISMA checklist, which ensures that all required parts have been described.[90] Keeping records of all decisions from the start in such a way that the PRISMA flowchart can be compiled helps with this process and offers a template to work from in a systematic way. As of 2018, PRISMA offers a template of the checklist specific for scoping reviews.[158]

If the analyses, writing the paper and peer review have taken many months or even years, it can be relevant to repeat the original search and scan for new data. That way newly included trials can be included and the results are more up to date.[129] Similarly, searching for SRs with the same research question can yield new publications. Discussing the findings of other SRs and the reasons for potential differences between SRs allows the readers to put the results into context in the wider literature landscape.

An essential part of reporting in any publication is the conflict of interest. Ideally, only objective investigators with no ties to any stakeholders would perform SRs,[8,51,81] but often content experts in their field have some link to commercial parties. In a study on the ties with stakeholders of SR authors, 29% of meta-analyses had authors who were employees of the drug manufacturer that was researched, and 79% had authors with some link (e.g. sponsorship or (former) employment). This is likely to consciously or subconsciously influence the assessment of the evidence and overall conclusion,[81] as research has shown that SRs or MA usually reach a conclusion that is favourable for the sponsor.[15,159]

Similarly, speciality and occupation of the authors influences the findings of SR,[122,159,160] as authors are more likely to be more critical of the methodology of trials that do not concur with their opinion than with those that do agree.[81] This further highlights the need for independent authors with no prior opinion or interest in the use of a treatment.

2.3 Errors in design and appraisal

This detailed description on how to perform an SR might make it seem like it is a cookbook exercise that can be implemented easily for any given question. There are many examples of bad methodology for SRs, though. Researchers search inappropriate or insufficient numbers of databases

to find trials,[141] oversee fundamental clinical insights or use inappropriate methods of integrating the data. In this paragraph, some clear examples will be mentioned.

2.3.1 Inclusions: Dieulafoy lesions

In their meta-analysis,[161] investigators compiled all forms of non-variceal upper gastrointestinal bleeding, including Dieulafoy lesions, despite the evidence that the bleeding rates are very different[162] and despite the known differences in response to specific treatments.[44] The conclusion was phrased as valid for all indications, leading practitioners to apply the findings to patients with indications other than those that are supported by the evidence.

2.3.2 Selection: Alcohol and all-cause mortality

A highly debated example in nutrition is the protective effect of alcohol on mortality. Several meta-analyses had found that moderate consumption, indicating one or two glasses of alcohol per day, was associated with a reduced risk compared to abstinence or heavy drinking for all cause mortality[163–168] or cardiovascular mortality.[169,170] One analysis found more than twenty diseases that are to benefit from moderate alcohol consumption (including deafness, the common cold, several psychiatric disorders, rheumatoid arthritis and asthma),[171] whereas another found an unlikely benefit of moderate drinking on alcoholic liver cirrhosis.[172] When plotting alcohol consumption against mortality risk, a J-shaped curve was typically found with the lowest risk around moderate consumption. This was considered convincing enough to recommend abstainers to start drinking moderately to improve their health.

Despite the difficulties in reliably establishing alcohol consumption, with issues such as recall bias, social desirability bias, compiling different consumption patterns to average consumption per week (seven drinks twice per week versus two drink every day), and fundamentals such as reliability of estimating volumes and percentages of drinks, the repeated finding of the J-curve made the findings credible.

The main concern over these findings, though, is the misclassifications of non-drinkers. Most meta-analyses do not control for the fact that those who previously consumed alcohol but stopped, or even alcohol-dependent subjects who stopped completely, will be classified as abstainers, skewing the risk of abstainers towards higher

risks. Several recent analyses[173–176] that controlled for this did not confirm the J-curve. Then there are 'sick quitters,' people who stop drinking once they get diagnosed with a severe ilness.[171,176] Additionally, moderate drinkers tend to have better socio-economic status, a healthier diet and more exercise than abstainers, factors which are not always accounted for.[171] Finally, as there are differences in the age at which diseases typically present, competing risks should to be taken into account.[176] This casts doubt on these findings, and on the methodology.

A large-scale RCT (MACH15, NCT03169530) to find a conclusive answer to this question was terminated early due to the influence of the funders (Anheuser-Busch InBev, Carlsberg, Diageo, Heineken, and Pernod Ricard).[177] For now, the health effects of drinking moderate amounts of alcohol will remain debated. One thing that is clear, though, is that the method used in the analysis significantly impacts the results.

2.3.3 Results: Cumulative PFS

An example highlighting the need for proper analysis is a recent review that added up median progression-free survival from separate clinical trials to come to "cumulative PFS,"[178] though medians cannot be added. Furthermore, median PFS was deduced from separate studies, with different inclusion criteria. The result is a highly unreliable "cumulative PFS," which could wrongly direct HCPs to the use of one treatment over another.

2.3.4 Lack of appraisal: Replication crisis

A clear example of what happens to a research field when SRs lack altogether is the replication crisis. As discussed in paragraph 2.2.5.1, many results are not confirmed when the research is replicated. This came to light when the open science collaboration tried to replicate 100 studies in psychological research that were published in high-ranking journals, and found that less than half of the results of these studies were replicated in follow-up research.[150]

There are several reasons for this, all of which could have been flagged by SRs. Because SRs are not as commonly used in psychological research as in more head-to-head EBM, this research field went on without the issues being flagged.

2.4 Tailoring SR methodology to research questions

Chapter 2 shows that there are many considerations when designing an SR. The databases that are used, the search terms that are chosen and the limits that are applied all impact which records are found. The method of calculating an overall effect is equally important. It also shows that there is still room for improvement in the methodology for SRs, even in the more mature and formal realm of EBM. Errors in the design can lead to wrong outcomes, with impacts on daily practice and real people, and can lead to unnecessary harm. As such, the careful design of SRs is important.

There are a couple of design choices that will be markedly different from EBM in our health-related research questions. First, the search methodology will be different, as the language is not as standardised as it is in EBM. As discussed in paragraph 2.2.3.2, PICO criteria are often used in search strategies for EBM, but this is not always applicable outside EBM. The key terms, such as drugs and diseases, are well-defined and indexed in thesauri, but other health-related questions cannot rely on this indexation as the research questions are often broader and less standardised.

Second, the choice of data integration methodology is not always as straightforward as in head-to-head clinical comparisons. Where an analysis of the pooled effect of a drug on overall survival seems straightforward, for other questions the scoring method requires thought and consideration.

Third, establishing bias and reliability of the findings is often harder. As the results of the selected sources are less standardised, assessing bias using the methods described in paragraph 2.2.7 is often not possible. For example, a discussion on how likely the effect of a publication bias is, will not be based on calculations, but on reasons.

To illustrate these considerations, the lessons from chapter 2 will be applied to specific research questions. In the next chapters, we will discuss the choices behind each methodology, and why it is appropriate to answer the question. We will also discuss the possible impact each design choice had on the outcome and the lessons learned from that.

The order of the chapters aligns with the design choices that are described above. In chapters 3, 4 and 5, we consider different options for the search methodology. In chapters 6 and 6 we discuss the ways to integrate data and calculate an effect. In chapter 6, we assess publication

bias in the findings, and consider how this affects the reliability of our conclusion. In chapter 7, we consider whether an SR is really the right method for our research question and consider an alternative approach. These case studies will allow us to answer our overall research question: *What choices and considerations can be made in the design of an SR for a health-related research question, and how do they impact the outcome and conclusion?*

3 Neurobiological correlates in forensic assessment: a systematic review

The first consideration when writing an SR in a health-related discipline outside EBM is the formulation of the search strategy. Drugs and diseases are indexed in thesauri, as discussed in paragraph 2.2.3.1. A thesaurus is a hierarchical reference structure that groups terms with similar meanings together, so that all papers can be tagged using a standardised language. These tags then allow users to search on these terms and find all papers that are related. So, for a typical SR in EBM, a reviewer would search for the disease terms and drugs that are of interest. For areas that are not as well-defined as a drug and a disease, a more tailored search needs to be designed.

This SR uses one possible search strategy. We used a sophisticated approach to find the right literature, selecting all the relevant terms in the thesauri of each search engine, tailored to the research question. We used this approach to still make use of the indexing, which would be more reliable than a free text search.

We wanted to include information from three search engines: Embase, PubMed and PsychInfo. They each had their own thesaurus, which meant we had to determine for each which terms we would include.

3.1 Context

The research objective of this SR was to provide an overview of all the current evidence about biological risk factors that predispose people to antisocial and violent behaviour, and to determine its value in forensic risk assessment. This was such a broad question that researching the primary literature only and assessing every study individually would be too time-consuming to reach a conclusion in a reasonable timeframe, so a mixed method was used.

First, an umbrella review was performed. Including previously published reviews allowed for a broad overview of the literature. The timeframe was 2000–2013, or 14 years. The benefit of this is that a large body of literature can be captured with relatively little effort, but the downside is that we had to trust in the judgement and methodology of

earlier authors to select the right papers, and that there was a delay in the research we included. That is why we performed another search.

Second, we performed a search for primary literature in a more recent timeframe: 2010–2013, or 4 years. This allowed us to include the insights of new research, not all of which had been critically assessed and included by authors of other SRs. This mixed method assured both a comprehensive and timely set of literature.

The search methodology was based on a selection of keywords in each thesaurus for PsycINFO, Embase and Pubmed (Thesaurus of Psychological Index Terms, Emtree headings, and MeSH, respectively). The terms concerned criminality, aggression, antisocial behaviour or psychopathy in combination with either neurosciences or genetics. The inclusion of specific terms was discussed by the authors to reach an agreement on the inclusion of the right keywords. This led to a search spanning 700 search terms.

After the searches had been performed in PsycINFO, Embase and Pubmed, and duplicates were removed, two independent authors selected the papers for inclusion in the review based on title and abstract. The differences in selection were discussed until a consensus was reached. The PRISMA reporting methodology was used to ensure quality of reporting.

3.2 Full text

A preliminary version of this chapter was published in PLoS One.

van der Gronde, T., Kempes, M., van El, C., Rinne, T., & Pieters, T. (2014). Neurobiological correlates in forensic assessment: a systematic review. *PloS One*, 9(10), e110672.

https://doi.org/10.1371/journal.pone.0110672

Neurobiological correlates in forensic assessment: a systematic review

Toon van der Gronde[1], Maaike Kempes[2], Carla van El[3], Thomas Rinne[2], Toine Pieters[1]

[1]*Department of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), and Freudenthal Institute, Utrecht University, Utrecht, the Netherlands.*

[2]*Netherlands Institute of Forensic Psychiatry and Psychology, Pieter Baan Center, Forensic Psychiatric Observation Clinic, Utrecht, the Netherlands*

[3]*Section Community Genetics, Department of Clinical Genetics and EMGO+, VU University Medical Centre, Amsterdam, the Netherlands*

3.2.1 Abstract

Background: With the increased knowledge of biological risk factors, interest in including this information in forensic assessments is growing. Currently, forensic assessments are predominantly focused on psychosocial factors. A better understanding of the neurobiology of violent criminal behaviour and biological risk factors could improve forensic assessments.

Objective: To provide an overview of the current evidence about biological risk factors that predispose people to antisocial and violent behaviour, and determine its usefulness in forensic assessment.

Methods: A systematic literature search was conducted using articles from PsycINFO, Embase and Pubmed published between 2000 and 2013.

Results: This review shows that much research on the relationship between genetic predisposition and neurobiological alterations with aggression is performed on psychiatric patients or normal populations. However, the number of studies comparing offenders is limited. There is still a great need to understand how genetic and neurobiological alterations and/or deficits are related to violent behaviour, specifically criminality. Most studies focus on only one of the genetic or neurobiological fields related to antisocial and/or violent behaviour. To reliably correlate the findings of these fields, a standardization of methodology is urgently needed.

Conclusion: Findings from the current review suggest that violent aggression, like all forms of human behaviour, both develops under

specific genetic and environmental conditions, and requires interplay between these conditions. Violence should be considered as the end product of a chain of life events, during which risks accumulate and potentially reinforce each other, displaying or triggering a specific situation. This systematic review did not find evidence of predispositions or neurobiological alterations that solely explain antisocial or violent behaviour. With better designed studies, more correlation between diverse fields, and more standardisation, it might be possible to elucidate underlying mechanisms. Thus, we advocate maintaining the current case-by-case differentiated approach to evidence-based forensic assessment.

3.2.2 Introduction

Violent crime is a complex problem without simple solutions. Given the prevalence of violent criminality in our society,[1–3] an understanding of the predictive and causal factors of violence is needed to improve assessment of criminal responsibility, risk assessment and management practices. What factors put individuals at risk for developing violent behaviour and committing a crime? What factors promote resiliency and protect individuals from re-offense? In the last 50 years, much has been learned about psychosocial risk factors that predispose people to violence.[4,5] However, psychosocial and biological causes of crime are inseparably entwined and are constantly interacting. Over the past two decades, on the tails of the genome project and a revolution in brain imaging, scientists across the world have tried to solve the enormous jigsaw puzzle of the biology of violent and criminal behaviour. These efforts have advanced our knowledge about and understanding of the biological factors and mechanisms involved in violent and criminal behaviour.[6,7] Nevertheless, forensic (risk) assessment is still mainly based on psychosocial risk factors.[8–10] The challenge now is to integrate these innovative neurobiological and genetic findings with current criminal assessment practices. When an individual suffers from a severe mental disorder that leads to a crime, it is generally agreed in most jurisdictions that he or she cannot be held criminally responsible and should be exempt from penal consequences.[7,11–13] Psychiatrists and psychologists are often called upon to render the expert opinions needed for legal determinations of criminal responsibility and risk for recidivism.[14,15]

Currently, forensic assessment is predominantly focused on psychosocial factors, however, till date, risk assessment instruments do not include biological risk factors.[8–10] We do know that psychosocial factors interact with biological factors in shaping (violent) behaviour.[6] Research shows that a small proportion of offenders, approximately 6%, account for the majority of all crimes[16] and that 5% of families account for more than 50% of all arrests.[17] With the increased knowledge of biological risk factors, interest is growing to include (more) information about biological risk factors in forensic assessments. In recent years, neuroscientific evidence, e.g. neurogenetics[18] and neuroimaging, has begun to be used to document a person's tendency towards aggression as was done in the case of the serial killers Brian Dugan[19] and Bradley Waldroup[20] and two recent murder cases in Italy.[11,20,21] This may yield several benefits. First, biological risk factors would lead to more objective measures of criminal responsibility or risk assessment of violent behaviour since they are thought to be less prone to manipulation. Second, assessing biological risk factors may reveal new information that could not be previously determined, such as assessing the possible role of a specific brain damage in criminal behaviour. Third, assessing information on biological risk factors would provide more information on the interaction between social and biological risk factors and their relationship with violent behaviour.

In summary, a better understanding of the neurobiology of violent criminal behaviour would help to provide insight into whether and how assessment of biological risk factors could improve forensic assessment. This systematic review aims to provide an overview of the current evidence about biological risk factors that predispose people to antisocial and violent behaviour and determine its usefulness in forensic assessment. As a framework to review the available literature, we adopt a biosocial model of violence as used by Raine.[6] Thus, we focus on evidence from genetics and interaction with pre- and post-natal environments, as well as related areas such as neuroanatomy, neuropsychology and neurology, neurophysiology, neurochemistry and endocrinology. This multi-disciplinary approach offers additional insight into the criminal mind and the underlying causes of violent behaviour to assist in forensic assessment.

Our hypothesis is that a general model that holds on a population level is, as of now, not evidence based. Forensic assessment remains to

be done on a case-by-case base. Though several pieces of knowledge can be connected, an overall picture that leads to a general understanding of criminal violence is not yet possible.

First, we present brain (dys-)functioning and behavioural effects with a special focus on brain anatomy and neurotransmitters. Subsequently, we review genetic and environmental influences on antisocial behaviour, as well as possible correlations with risk factors. The relevant models and theories for each section will be discussed, including supporting evidence for each one. Finally, we consider the possible implications for forensic assessment and address research challenges.

3.2.3 Method

3.2.3.1 Data sources

Two literature searches were conducted in the electronic databases of PsycINFO, Embase and Pubmed. The first search concerned reviews and/or meta-analyses published between 2000 and 2013; the second search retrieved empirical research published between 2010 and 2013. We selected publications by using a query based on keywords concerning criminality, aggression, antisocial behaviour or psychopathy in combination with either neurosciences or genetics. The exact query is provided in the appendix.

3.2.3.2 Inclusion

In both searches we included publications using the following criteria: a) published in peer-reviewed journals, b) written in English, and c) included offender populations. We excluded papers that were: a) case reports, books, conference abstracts, letters, b) written in languages other than English, c) published before the dates mentioned, d) animal studies, or e) concerned only paedophilia/paraphilia due to the likeliness of these being caused by mechanisms other than violent criminality.

3.2.3.3 Selection

All selected publications were assessed for relevance based on both title and abstract. The articles were judged for inclusion by two independent researchers based on content. The reviews and/or meta-analyses were divided into two groups. The first group of articles related to a criminal or forensic context. The second group of articles addressed types of behaviour, i.e. physical aggression and violence that are most relevant to the criminal justice system in terms of personal damage for

the victim and serious legal consequences for the perpetrator. Twenty articles that were deemed essential but were not found in the systematic search were added. This is shown in figure 1.

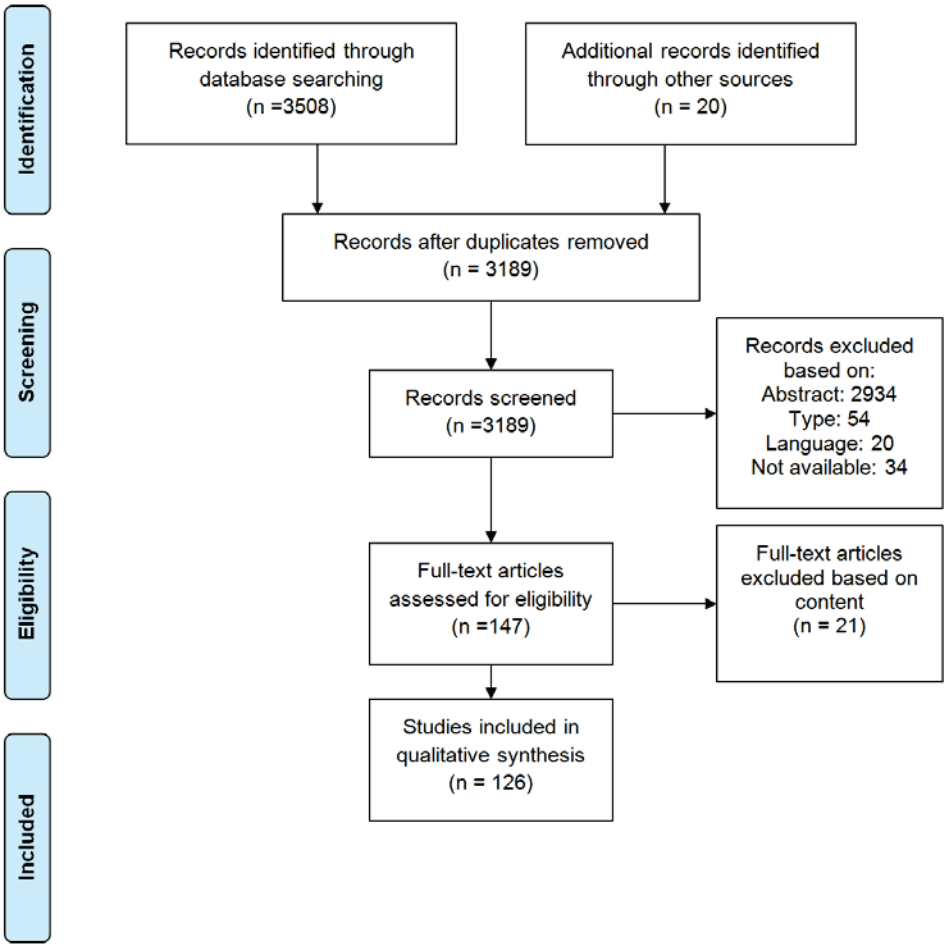

Figure 1. Selection of publications. Out of the 3508 found articles, 126 were used for this article.

3.2.4 Results

3.2.4.1 Brain and criminology

Various disciplines study brain functioning and the effect dysfunction has on behaviour, but each has a distinct perspective and aim. The following section aims to correlate the findings across these research fields to facilitate a more thorough understanding of the possible

relationships between criminal behaviour, neuroanatomy, brain biochemistry and neuropsychology.

We first present neuroanatomy, including the morphological structures of the brain that are found to be relevant in imaging studies. Next, we discuss forensic neuropsychology – this encompasses the integration of psychological findings with neurology by performing tests that specifically target an area of the brain.[9,22] Finally, an overview of the related neurotransmitters and hormones is presented. A schematic summary of the evidence from the different research fields can be found in a series of six tables.

3.2.4.2 Neuroanatomy

One of the challenges of neurocriminology is to trace biological markers of sociopathy with brain imaging research. Brain-imaging techniques identify physical deformities and functional abnormalities that may predispose some individuals to violence. This has led to theories of neuroanatomical deviations and criminal behaviour. In the following paragraphs, these theories and evidence for them are discussed. However, first we provide a brief overview of the most important structures mentioned in research on violent criminal behaviour and the relationships between these structures. (See table 1)

An important factor in violent criminal behaviour is emotion, or the lack of it. In general, the amygdala is involved in emotions,[23–27] particularly negative emotions,[27–29] and the recognition of fear [26,29,30]. The regulation of emotions is thought to be dependent mostly on the prefrontal cortex,[1,3,13,27,31–36] which is rich in 5-hydroxytryptamine (5-HT) type 2-receptors.[2] The prefrontal cortex inhibits the amygdala, as well as other limbic system regions, like the hippocampus (involved in memory[14]), hypothalamus, anterior cingulate cortex, insular cortex, ventral striatum, and some structures connected to those regions.[1,2,31,36] The prefrontal cortex also receives information from the rest of the cortex and the limbic system.[1,37]

Another important aspect in the instigation of behaviour is processing social information in one's environment. Signals that can indicate a threat, like posture, facial expression or screaming, are directed to the lateral nucleus of the amygdala.[2,26,31,38–40] The signals project to the basal nuclei, where they are integrated with

perceptual information originating from the orbitofrontal cortex.[2,24,41] This can lead to a behavioural response via the central nucleus, the hypothalamus and brainstem.[2] Thus, the orbitofrontal cortex is thought to integrate the cognitive activity of the total prefrontal cortex into the emotional limbic system.[25–27,37] In this manner, the prefrontal cortex restricts impulsive, disinhibited behaviour and volatile emotions.[2,25,26,31]

Dysfunction or neuroanatomical deviations of one or several of the above-described structures have been shown to be related to violent criminal behaviour. However, before we describe these studies, it is important to distinguish between two types of violent aggressive behaviour. Reactive,[42–48] emotional[45] or impulsive[2] aggression is a reaction to events, often driven by emotion. Instrumental[42–45,47,48] or premeditated[2] aggression, however, is cold and calculated. Since reactive behaviour is influenced primarily by emotions, and the instrumental is not, distinct brain areas are expected to be involved in the various forms of aggression.[2,42,44,46,48]

3.2.4.2.1 Prefrontal cortex

The frontal lobe dysfunction theory states that violent and reactive aggression is a consequence of deficits in the frontal brain, mainly the prefrontal cortex.[44,44,49,50] Supporting evidence for this theory comes from various research areas.

First, research conducted with PET shows reduced functioning of the prefrontal cortex in offenders,[6,33,37,51,52] and reduced activity in violent patients. [36] The association is stronger in murderers with a benign social background, than in those with a bad home background,[6] as expected based on the social push theory (which will be discussed later on). Frontal abnormalities have also been found using EEGs on aggressive subjects in various populations (like violent criminals [53]), with PET in forensic psychiatric patients[31,53] and in an MRI study of antisocial patients.[3,53] One study distinguished between the various regions of the prefrontal cortex, and specifically found reduced activity in the ventrolateral part of the prefrontal cortex – relevant for social behaviour[54] – in aggressive subjects.[44]

Second, studies on brain metabolism show that in general, a reduced glucose metabolism in the prefrontal regions can be found in violent offenders.[32,55] Specifically, it was found that murderers[56]

and violent psychiatric patients[31,33] have a lower prefrontal cortex metabolism than controls. This finding has been replicated in impulsive murderers for whom a rise in the metabolism of subcortical regions was also found, as expected, since inhibition by the prefrontal cortex is reduced.[2] However, murderers in that same group who planned their crimes did not have a lowered metabolism in their prefrontal cortex.[2] When the findings for predatory and affective murderers were separated, it was clear that affective murderers had lower prefrontal metabolic activity than predatory murderers, who resembled controls.[32,53]

Third, a study examining N-acetyl aspartate, considered a marker of neuronal integrity,[3,57] showed that violent patients had less N-acetyl aspartate in their prefrontal cortex compared to controls.[3,32] More importantly, the frequency of violence was inversely correlated with the levels of N-acetyl aspartate.[3] A lower phosphate metabolism was also found in the prefrontal cortex of the violent patients.[3,32]

In general, deficits in the prefrontal area, mainly the ventromedial part, have been found to be related to poor control of reactive violence.[2,31,45,46,48] Instrumental violence, on the other hand, is thought to be associated with the dysfunction of both the ventromedial prefrontal cortex and the amygdala.[46] In addition, individual differences appear to be important. Personal differences in the ability to modulate emotions have been shown to be linked to prefrontal activation.[2] This might be relevant in understanding vulnerability to violence and aggression.[2]

3.2.4.2.2 Amygdala

The amygdala is comprised of thirteen nuclei, together forming one structure.[54] The integrated emotion systems model hypothesises that deviant social behaviour, such as violence, is the result of inhibited emotional development caused by amygdala dysfunction.[28,35,58–60] This is supported by a study in murderers , that showed a lowered amygdala activity, compared to age- and sex-matched controls.[42]

Moreover, the integrated emotion systems model states that amygdala damage leads to impaired interpretation of emotions.[23,61] This results in diminished empathy, failure to recognise fearful expressions and impaired passive avoidance learning, all of which have been documented in psychopaths.[25,58]

These studies lead to the conclusion that prefrontal cortex function[3,53] or size[1] and amygdala function[3] are related to violent aggression. It is possible that both the frontal lobe dysfunction theory and the integrated emotion systems model are true, and reinforce each other. It is also possible that the former theory explains reactive aggression and the latter explains instrumental aggression.

3.2.4.2.3 Hippocampus, temporal lobe, anterior cingulate cortex

In addition to the amygdala and prefrontal cortex, other brain structures have also been shown to differ in criminal subjects compared to the general population. The hippocampus is part of the limbic system and is involved in memory. The hippocampus has been shown to function abnormally in violent offenders[3] and subjects who commit murder[6,55] and is structurally distinct in psychopaths.[6]

The temporal lobe contains the hippocampus and plays a key role in the formation of explicit long-term memory that is modulated by the amygdala. A 20% reduction of the temporal lobe was found in aggressive psychopaths,[62] and functional abnormalities of the temporal lobe were found in violent psychiatric patients.[33] Sexual offending might also originate in the temporal lobe.[37,63,64] However, this goes beyond the scope of this review.

The anterior cingulate cortex is a limbic region involved in response selection, behavioural regulation, inhibition,[65,66] and empathy.[26] Interestingly, anterior cingulate cortex hemodynamic activity predicts re-arrest; higher activity leads to better inhibitory control and recurrence rates that are half of those with low activity.[65] If this correlation is replicated, it could possibly lead to a better tool for risk analysis in combination with known psychosocial risk factors, as scepticism remains regarding the sensitivity and specificity of emerging neurobiological markers as independent tools. [61]

Table 1. Overview of the evidence for involved brain areas. PET: Positron Emission Tomography, EEG: Electroencephalography, (f)MRI: (functional) Magnetic Resonance Imaging, SPECT: Single-photon emission computed tomography, ERP: event-related potentials, CT: computed tomography

Brain region	Population	Method	Outcomes	Reference
Prefrontal cortex	Offenders	PET	Reduced functioning	[6,33,37, 51,52]
	Violent patients	PET	Reduced activity	[36]
	Aggressive subjects	EEG	EEG abnormalities	[53]
	Forensic psychiatric patients	PET	Decreased blood flow or metabolism	[31,53]
	Antisocial patients	MRI	Reduced frontal grey volume	[3,53]
	Aggressive subjects	PET	Reduced ventrolateral activity	[44]
	Violent offenders	PET	Reduced glucose metabolism	[32,55]
	Murderers	PET	Reduced glucose metabolism	[56]
	Violent psychiatric patients	PET	Reduced glucose metabolism	[31,33]
	Impulsive murderers	PET	Reduced glucose metabolism	[2]
	Violent patients	MRI	Less N-acetyl aspartate	[3,32]
	Violent patients	MRI	Lower phosphate metabolism	[3,32]
Amygdala	Murderers	PET	Lowered activity	[42]
	Psychopaths	fMRI	Reduced activation	[3]
Hippocampus	Violent offenders	SPECT	Low resting blood flow	[3]
	Violent offenders	PET	Reduced metabolism	[3]
	Murderers	PET	Abnormal functioning	[6,55]
	Psychopaths	MRI	Structurally distinct	[6]
Temporal lobe	Impulsive-aggressive Personality-disordered	MRI	20% reduction	[62]
Anterior cingulate cortex	Offenders	fMRI	Hemodynamic activity predicts re-arrest	[65]
Hemispheres	Several populations	Several techniques	Left dorsolateral prefrontal cortex deficits	[49]

	Several populations	Several techniques	The right orbitofrontal cortex and anterior cingulate cortex right	[49]
	Affective and predatory murderers	PET	High right hemisphere subcortical functioning	[31]
	Affective murderers	PET	Low left, high right prefrontal functioning	[31]
	Violent psychiatric patients	SPECT	Increased or abnormal left limbic activity	[33]
	Offenders who were victims of child abuse	fMRI	Reduced right temporal cortex functioning	[6]
	Psychopathic patients	MRI	Reduced right temporal cortex volume	[52]
	Incarcerated psychopaths	fMRI	Dysfunction in the right hemisphere during abstract processing	[56]
	Antisocial and violent subjects	fMRI, CT, EEG, ERP	Poor right hemisphere functioning	[32]

3.2.4.2.4 Hemispheres

Apart from findings concerning specific brain areas, much research is directed at structural dysfunction in the left or right hemispheres. The left Hemisphere Activation Hypothesis states that psychopaths have problems shifting from left hemisphere activity to right hemisphere activity, and specifically processing information in the left hemisphere.[49] Support for this hypothesis is offered by deficits in the left dorsolateral prefrontal cortex, [49] which is associated with attentional control.[27] This makes sustaining attention in the left hemisphere more difficult.[49] Deficits in the right orbitofrontal cortex and anterior cingulate cortex support this since they are involved in the ability to change to right hemisphere activity.[49]

In one study, both affective and predatory murderers had higher subcortical right hemisphere functioning than controls, but affective murderers also had lower left and higher right prefrontal functioning.[31] In another study, violent psychiatric patients were found to have increased or abnormal left limbic activity.[33] Offenders who were victims of child abuse have been shown to have reduced right temporal cortex functioning,[6] which is associated with conduct

disorder.[26] Reduced volume of the right temporal cortex has been found in psychopathic patients.[52]

General deficits in the right hemisphere have been proposed as well. During abstract processing, which is thought to be based in the right temporal lobe, a dysfunction in the right hemisphere was found in incarcerated psychopaths.[56] In antisocial and violent populations, poor right hemisphere functioning has also been observed.[32]

Overall, several brain areas appear to be deficient in violent individuals. An overview of findings is given in table 1. However, it is unknown whether these deficits always result in violent behaviour per se, since this was not investigated in the studies mentioned. The evidence that brain deficits are related to violent behaviour is mainly based on case reports such as the one on Phineas Gage (this case will be discussed in more detail later on), which have been shown to be untrue or only partly true.[21] Moreover, case reports that show no violent behaviour in persons with brain deficits are also available.[21] Therefore, more research should be done on how brain deficits or alterations are related to the actual instigation of violent behaviour. In this respect, it may be more fruitful to examine how altered brain *function* is related to violent behaviour.

3.2.4.3 Neuropsychology and neurology

The brain areas described above have distinct functions. In the following paragraphs, we describe additional evidence for a relationship between altered brain *function* and the propensity for violent behaviour. A schematic overview of the evidence is given in table 2.

3.2.4.3.1 Executive functioning

As shown above, murderers' prefrontal cortices often seem to be an affected brain area. Numerous studies have demonstrated that the prefrontal cortex is important in executive functioning [1,6,32,37,43,50,53,60] (organising cognitive processes), to attain a future goal. [1,67] Executive functioning includes attention control, behavioural flexibility, working memory, self-awareness, abstract decision making and planning. [6,14,43,63,68] Executive functioning is required for much complex behaviour, such as social functioning and managing competing interests,[67] and can be measured using neuropsychological tests. [53,68]

Lesions in or dysfunction of the prefrontal cortex[37] or the frontal lobe in general[51] lead to impaired executive functioning. Impaired executive functioning is associated with antisocial[6,68] and aggressive[51,53] behaviour. More importantly, low executive functioning can predict aggressive behaviour in boys with a paternal history of substrance abuse.[53] This could possibly help determine the risk for recidivism in general.[53]

3.2.4.3.2 Inhibition

Another structure that has been shown to be significant in violent behaviour is the anterior cingulate cortex. The anterior cingulate cortex[65] and the serotonergic neurons in the prefrontal cortex are thought to be important for behavioural inhibition.[37] Indeed, prefrontal damage, especially orbital damage,[8,37,68] does lead to lower inhibition and pseudopsychopathic behaviour.[8,69]

3.2.4.3.3 Empathy

Several brain regions are involved in the instigation of empathy. Lesions in the orbitofrontal cortex[53], prefrontal cortex,[41] amygdala[41,54] or anterior cingulated cortex[41] are related to a lack of empathy. A lack of empathy is indeed often found in offenders.[37,70] These abnormalities in the emotional regulation circuitry are thought to lead to reactive aggression and the violence seen in these individuals.[2]

3.2.4.3.4 Psychophysiology

Differences in physiology between the offending and general population have been found, long before brain-imaging techniques existed. In this respect, one of the most replicated observations in aggressive anti-socials and psychopaths is low autonomic arousal in rest, measured by resting heart rate and skin conductance.[41,45,52,55]

It is likely that low autonomic arousal is related to or a result of anatomical and functional deviances in violent offenders. For example, it is proposed that reduced noradrenergic functioning and reduced right hemisphere functioning would explain the low autonomic arousal found in violent criminals.[32] Moreover, autonomic arousal is also controlled by the amygdala,[28] which has been found to be less functional in murderers. Therefore, low arousal may be a marker for amygdala dysfunction.[39]

Although the origin of low autonomic arousal is of interest, forensic risk assessment would particularly benefit from knowledge

about how low autonomic arousal may be related to or predict (violent) criminal behaviour. Several theories are of interest.

First is the fearlessness theory, which states that low levels of arousal are a marker for low levels of fear.[3,6,32,52,71] Fearlessness predisposes a person to criminal behaviour, because criminality requires low fear levels.[6,32,71] Also, the effectiveness of learning through conditioning is diminished by less anticipatory fear for punishment,[32,44,52,58,62,71,72] leading to impaired socialisation.[6,42,58,71] Second, the stimulation-seeking theory states that low arousal will make subjects seek more exciting, possibly criminal activities, trying to relieve their boredom.[32,42,52,55,58,71,72]

In accordance with the notion that low autonomic arousal may not only be related to but may also predict criminal behaviour, several studies have shown a link with future criminal offenses,[32,52,55,72,73] aggression[6,32,55,62,71,72] – especially instrumental aggression[42] – and antisocial behaviour.[6,32,41,55,58,62] It has even been shown that low autonomic arousal is predictive of children growing up to become offenders.[32,55,66] In one study, aggressive children had lower heart rates than nonaggressive children ($p<0.001$), and children with lower heart rates were rated as aggressive more often than those with high heart rates ($p<0.003$).[71] Therefore, autonomic arousal may be an interesting marker to improve risk assessment in future. A downside to using autonomic arousal as a marker is that as of yet it is unknown what the cut-off point for increased risk would be.

Table 2. Overview of the evidence for brain functions

Brain function	**Population**	**Method**	**Outcomes**	**Reference**
Executive functioning	Boys with a paternal history of substance abuse	Neuropsychological tests	Low executive functioning can predict aggressive behaviour	[53]
Psycho-physiology	Aggressive anti-socials and psychopaths	Resting heart rate and skin conductance	Low autonomic arousal in rest	[41,45, 52,55]
	Children	Resting heart rate and skin conductance	Low autonomic arousal is predictive for becoming offenders	[32,55, 66]

3.2.4.3.5 Neurotransmitters, hormones, and toxins

Neurotransmitters and some hormones are important for communication between neurons in the brain and thus, they are of importance in the instigation of behaviour. Therefore, researchers have sought the origin of criminal behaviour in a disturbed balance between some of these neurotransmitters or hormones. Since toxins influence the levels of these neurotransmitters, they could also be of significance. A schematic overview of the evidence is provided in table 3.

3.2.4.3.6 Serotonin

One of the most replicated findings is the relationship between serotonin and aggression. Numerous studies have shown that low levels of serotonin are associated with both reactive and instrumental aggression [2,7,8,10,25,31,34,41,43,45,51,66,74–78] and impulsivity. [7,10,22,25,31,34,37,43,51,72,77,79] In addition, low serotonin levels [31,55] and reduced levels of 5-hydroxyindoleacetic acid, a serotonin metabolite, have been found in aggressive or violent populations. [2,31,41,43,52,74] Furthermore, a negative correlation between the serotonin 5-HT1A receptor and aggressive behaviour has been established. [79] In impulsive aggressive subjects, reduced serotonin transporter availability was found in the anterior cingulated cortex. [78] Moreover, one study showed that low levels of serotonin predicts recidivism. [7] 5-hydroxyindoleacetic acid levels have been found to predict aggression two to three years in the future in boys with conduct-disorder and recidivists. [2] Antidepressant drugs that act on serotonin, like SSRI's that cause serotonin levels to go up, can reduce violent behaviour in some individuals, [10,33,45,80,81] especially those with high impulsive aggressiveness. [48] Although the above-mentioned studies show that aggression and violence are related to low levels of serotonin, other results seem to indicate the opposite. Metabolic enzymes such as monoamine oxidase A (MAO-A) also contribute to aggression because they function to alter neurotransmitter levels. Since MAO-A catalyses the deamination of serotonin, reduced MAO-A activity will lead to higher levels of this neurotransmitter. [57,66,74,82,83] However, MAO-A deficiency, resulting in higher levels of serotonin, has been shown to increase reactive aggression [20,74,84] and low activity to increase criminal behaviour. [7,84] This is called the serotonin paradox. One author argues that the change in behaviour due to MAO-A is actually a consequence of secondary effects, and cannot be explained by its effect on neurotransmitters alone. [74] Taken together, the above-

described results do show that the relationship between serotonin and aggressive or violent behaviour is more complicated than is sometimes presented in the courtroom. [81] An individual risk-assessment on the basis of serotonin levels is not supported by evidence.

3.2.4.3.7 Noradrenalin

Although the relationship between serotonin and aggression and violent behaviour seems strong, there is also evidence that other neurotransmitters are involved. For example, noradrenalin levels, a neurotransmitter involved in the inhibition of memory storage and experiences,[48] in plasma and cerebrospinal fluid are positively correlated with impulsivity [31] and affective aggression. [45] This does not provide much information about the exact site of noradrenalin release, but makes drugs counteracting noradrenergic function interesting for preventing aggressive behaviour. [31,45]

3.2.4.3.8 Dopamine

Dopamine levels, a neurotransmitter important for rewards, delayed rewards and risk taking, [85] have been correlated with violent [86–88] and antisocial behaviour [82,89] and sensation seeking.[85,90] Activation of dopamine receptors, especially the D_2, [88] D_3 [7] and D_4 [87] receptors, are related to aggressive impulses, [2,25,45] and regulated by serotonin. [25] D_2 receptor agonists have successfully been used to treat aggression in some patient groups, especially those who are psychotic. [48,76]

3.2.4.3.9 GABA

Finally, another neurotransmitter, γ-aminobutyric acid (GABA), also seems to inhibit aggression. [8,45,48] Indeed benzodiazepines, substances that enhance GABA signalling, are effective in reducing aggression in humans, [45,91] though in specific subsets it increased aggressive behaviour. [48]

In summary although evidence for the effects of the serotonin system on violent aggression is strongest, several other neurotransmitters seem to affect aggression and violence. To make the situation even more complicated, the neurotransmitter system also has interactions with other systems in the body, such as the endocrine system.

3.2.4.3.10 Hormones

One of the most studied relationships is that between the stress system and aggression. Since the prefrontal cortex contains some of the highest levels of cortisol receptors of the primate brain, low levels of stress hormone will alter the turnover of various neurotransmitters. [1] Adrenocorticotropic hormone (ACTH) is produced when cortisol is suppressed, and it increases serotonin metabolism. [31] This results in lower serotonin levels. [31] Cortisol itself seems to be inversely correlated to levels of serotonin. [92] As such, low cortisol levels were associated with sensation seeking [41] and decreased sensitivity to punishment[39], but also with aggressive behaviour in boys, [10,41,43,55,93] adolescents [39,55,93] and adults [33,72,73,93]. However, similar to the serotonin system, these relationships are not unequivocal since high levels of stress hormone have also been found to be related to aggressive behaviour. [94] The key seems to be that the production of cortisol is deregulated.

A second important hormone in relation to violent aggression is testosterone. Plasma testosterone levels have been associated with childhood and adult delinquency, [6,93] antisocial behaviour, [64] aggression [2,6,10,33,55,66,72,93,95] and dominance, [25] but a correlation with social success has also been suggested. [93] These correlations have not always been well replicated. [93] The effect of testosterone on aggression is not visible in young children, possibly because aggression in childhood does not increase dominance as it does in adulthood. [6] In 9-11 year old boys, the association between testosterone and aggression has been documented. [55]

During development, testosterone induces or inhibits cell death, guiding the brain to typical male pathways. [43,74,93] Later, it stimulates neural pathways associated with aggression. [74] Testosterone receptors have in fact been found throughout the limbic system. [43] The association between testosterone and aggression has been confirmed by users of anabolic steroids. [74] In tests, testosterone injections led to a shift in sensitivity from punishment to reward. [39] Hypogonadal adolescents receiving testosterone became more aggressive physically, but not verbally. [93] This could be explained by changes in musculature as well. [93]

Testosterone and cortisol inhibit each other's production. [25,39,58] This means it is possible that the findings of the effects of

cortisol are actually due to testosterone, or the other way around. The triple balance of emotion model states that the hyposensitivity for punishment and the hypersensitivity for reward found in psychopaths [40] could be explained by a high testosterone-to-cortisol ratio. [25,39,52,58] Indeed, testosterone increases sensitivity to reward,[25,39] and low cortisol stimulates the hypothalamic–pituitary–gonadal axis, reducing sensitivity to fear. [39,58]

A third group of hormones related to antisocial behaviour are the thyroid hormones. T_3 and T_4 have been related to antisocial behaviour. [41] T_3 has also been specifically linked to recidivism. [41]

Table 3. Overview of the evidence for involved neurotransmitters and hormones. GABA: γ-aminobutyric acid, T3: triodothyronine

Neurotransmitters and hormones	Population	Method	Outcomes	Reference
Serotonin	Several	Several	Low levels of serotonin are associated with both reactive and instrumental aggression	[2,7,8,10, 25,31,34, 41,43,45, 51,66,74–78]
	Several	Several	Low levels of serotonin are associated with impulsivity	[7,10,22, 25,31,34, 37,43,51, 72,77,79]
	Boys with conduct-disorder and recidivists	5-hydroxyindoleacetic acid level measurement	Predict aggression two to three years in the future	[2]
Noradrenalin	Humans	Plasma and cerebrospinal fluid measurements, report scale	Noradrenalin is positively correlated with impulsivity	[31]
	Humans	Drug administering	Increases in affective aggression when noradrenalin is elevated	[45]
Dopamine	Humans, offenders	Gene expression	Activation of D2, D3 and D4-receptors are related to	[7,87,88]

			aggressive impulses	
GABA	Humans	Benzodiazepine use	Benzodiazepines, are effective in reducing aggression	[45,91]
Cortisol	Boys, adolescents and adults	Saliva measurement	Low cortisol levels were associated with aggressive behaviour	[10,33,39, 41,43,55, 55,72,73, 93,93,93]
Testosterone	Children, adults	Plasma testosterone	Delinquency	[6,93]
	Offenders	Saliva measurement	Antisocial behaviour	[64]
	Males	Plasma testosterone	Aggression	[2,6,10,33, 55,66,72, 93,95]
	Several	Several	Dominance	[25]
	Hypogonadal adolescents	Testosterone administration	More physical aggression	[93]
Thyroid hormones	Delinquent boys	Serum levels	Relationship between T3 and antisocial behaviour	[41]
	Former juvenile delinquents	Serum levels	T3 levels correlate with persistent criminal behaviour	[41]

3.2.4.3.11 Other substances

Both the endocrine and neurotransmitter system are influenced by substances in the body other than hormones. For example, the connection between alcohol and violence is well documented. [34,91,96] Over half of all violent crimes occur under the influence of alcohol. [91] Alcohol's mechanism of action is thought to be dependent on the function of $GABA_A$, [48] 5-HT and N-methyl-D-aspartate receptor (NMDA)-receptors.[91] A lowering of tryptophan, thought to be parallel to the level of brain serotonin, has been documented two hours after alcohol consumption in a normal population. [34]

In addition, alcohol inhibits the capacity of the prefrontal cortex, leading to impaired executive functioning. [1,97] This makes it a disinhibiting factor, leading to acting out what was previously inhibited.

[10,70,92,97] Alcohol is an aggravating factor in domestic violence, [33,51] and increases the chances of committing homicide. [98]

Abuse of other substances also increases risk of violence. [4,8,33,50,51,99,100] Cocaine for example enhances dopamine signalling, [1] and decreases the capacity to control impulses. [8,70]

Another substance that may affect neurotransmitter levels is cholesterol. Low cholesterol has been linked to aggression. [2,55] In community samples of psychiatric patients or criminal offenders with low cholesterol levels, an increase in violence was found. [55] A possible explanation for this observation is that low cholesterol leads to lower serotonin levels. [55]

To summarize, the various neurotransmitter systems in the central nervous system have complex interactions with each other and with other systems in the body such as hormones and toxins. This makes it hard to understand how aggression and violence are regulated in individuals.

3.2.4.4 Genetic and environmental influences

3.2.4.4.1 Genetics, GxE

Genetic influences on antisocial and aggressive behaviour have been documented in literature. [6,17,41,61,66,75,101–105] Given the influence that neurotransmitters and hormones have on aggression, a genetic basis of violence can be expected in related genes. [76,106,107] For example, the influence of serotonin transporters [18,41,108] and receptors, [61] tryptophan hydroxylase, [2,77] MAO-A, [83,106] catechol-O-methyltransferase, [18,77,89] dopamine receptors, [17,41,86–88,104,107] the androgen receptor [64,95,109] and the corticotrophin releasing hormone receptor [110] have been mentioned. However, in a meta-analysis including these genes, no single gene was significantly correlated with aggression. [76] Genes can still be used to have a better understanding of aggression, but not for risk assessments or to determine criminal responsibility. [76] A schematic overview of the evidence is given in table 4.

Gene-gene interactions can also be expected to occur. [102,107] Given the complex interplay of neurotransmitters, the effects of genetic polymorphisms can be corrected or aggravated by other genetic polymorphisms. [102,107]

However, aggressive behaviour is not caused by genetics alone. The 'social push'-theory states that genes need a particular social environment to result in specific behaviour. [6] Antisocial personalities, for example, develop due to biological factors, but lead to antisocial behaviour more often if the social situation predisposes, or pushes the individual to that behaviour. [42,56] On the other hand, if the social environment does not require antisocial behaviour to achieve what is wanted, antisocial behaviour might not develop despite an unfortunate biological background. [42,52]

This also means that the correlation between antisocial behaviour and biological risk factors becomes weaker in cases of poor social backgrounds, like a broken home. [42] This is because when the environment does not push an individual towards a negative behaviour (such as someone who has been reared in a benign social environment), but the antisocial behaviour comes to expression anyway, genetic factors have played a larger role in the instigation of the antisocial behaviour. [73] When the environment pushes too hard, like in very poor social backgrounds, every individual is influenced, resulting in a weak correlation between genetic factors and the actual behaviour. [42]

To study whether it is genetic makeup or the environment that causes specific behaviour, twin and adoption studies are often used. [40,45,69,82,103,105,109,111–113] This is because monozygotic twins share identical genetic material, and dizygotic twins share on average 50% of their dissenting genetics, [40,82,101,103,112,114,115], but both share an environment.[103] Subtracting the differences between these groups allows estimations of the contribution of environmental versus genetic factors when behavioural differences are measured. [101,103,114,115] Twin studies have, for example, shown the relevance of both genetics and environment for the development of antisocial behaviour, violence and aggression. [55,111]

The adoption method compares the correlation between adopted children and their adopting parents with the correlation between adopted children and their biological parents.[101] This also results in an estimate of the contribution of genetic and environmental factors. [101]

One of the most studied genes in research on gene-environment interactions is MAO-A (see table 4). A MAO-A deficiency has been shown to increase reactive aggression, [20,74,84] and its low activity increases

criminal [7,84] and antisocial behaviour. However, this last result was found especially in males when the subject had also suffered from childhood maltreatment. [10,20,76,86,106,114] In criminal settings, those who had a promoter sequence resulting in low MAO-A activity and who had been maltreated as a child were overrepresented. [20,116] Both examples illustrate that the effect of MAO-A is dependent on environmental factors, so the environment and genetics interact. [57,108,116]

Gene effects rarely influence behaviour directly. MAO-A, for example, may have a role in the difference between male and female levels of violence since the MAO-A gene is encoded on the X chromosome. [7,96,106,112,113] The documented correlation between high testosterone levels and low MAO-A activity, and resulting aggression, supports the hypothesis of further testosterone-induced suppression of the MAO-A gene. [113] The promoter region of the MAO-A gene does in fact contain glucocorticoid/testosterone response elements. [20] Testosterone competes with cortisol for binding, but leads to less transcription than cortisol binding does. [20]

Males have been found to have less connectivity between the orbitofrontal cortex and the amygdala, [113] lower functional connectivity between the ventromedial prefrontal cortex and the amygdala, [20] lower orbitofrontal activity [113] lower cingulate cortex activation, [113] and a larger amygdala. [117] This does not explain, however, the difference between male and female proclivity for violence. This example makes clear that the change in behaviour due to the MAO-A gene is actually a consequence of secondary effects, and cannot be explained by its direct effect on neurotransmitters alone. [74]

Table 4. Overview of the evidence for genetic influences. For a more complete overview, see Vassos' review [76].

Genetic influences	Population	Method	Outcomes	Reference
MAO-A	Subjects with childhood maltreatment	Genetic testing	Correlation between low activity and antisocial behaviour	[10,20,76,86, 106,114]

3.2.4.4.2 Prenatal environmental factors

The prenatal period is an important time for development of the brain and influences function and the way actual behaviour is instigated later on. Exposure to several addictive substances used by the mother during

this period influences brain development.[6,32,69] A schematic overview of the evidence is given in table 5.

3.2.4.4.2.1 Substance exposure

Prenatal alcohol exposure can cause structural deficits in the corpus callosum,[55,67] and cerebellum in the infant.[67] It also impairs the infant's memory[67] and executive functioning[67] and lowers IQ.[55,67] Though the physical signs diminish in adolescence, the neuroanatomical differences remain.[67] These changes may explain why it is also found that prenatal alcohol exposure increases the risk for conduct disorder.[6]

For nicotine a dose-response relationship between the number of cigarettes smoked during pregnancy and violence has been found.[6,55] Prenatal nicotine[6,32,55] and carbon monoxide[32,55] exposure is thought to disrupt the development of the noradrenergic system,[6] possibly via enhancing the muscarinic 2 (M_2)-receptor,[32] leading to diminished sympathetic nervous system activity. This could explain the observation of low autonomic arousal in violent and antisocial individuals and criminals.[6,32]

Prenatal cocaine exposure is also associated with increased delinquency, but these results are debated.[55]

3.2.4.4.2.2 Nutrition

Apart from addictive substances, basic nutrition during pregnancy influences the development of the baby and later behaviour. Like the well-known effects of folic acid on preventing spina bifida, other nutrients influence the development of aggressive behaviour.

Women who suffered nutritional deficits during the first and second trimester of their pregnancy gave birth to children who had antisocial personality disorder more often than the general population in two studies.[32,66] In addition, heavy metals like copper have also been shown to influence later behaviour. High copper in the neonatal brain is associated with abnormalities in the hippocampus,[55] which is associated with violence. A low zinc to copper ratio was found in males with a history of assaultive behaviour.[32]

In addition to the use of addictive substances or nutrition by a mother, birth complications are also a risk factor for prenatal development. Both of these factors could be seen as markers, although the exact mechanisms of these factors are not clear.[43] For example, a

significant interaction between maternal smoking and delivery complications has been documented.[43]

3.2.4.4.2.3 Birth complications

Birth complications such as anoxia, preeclampsia and forceps delivery lead to increased risk for antisocial and criminal behaviour through brain dysfunction.[6,32] The hippocampus is particularly susceptible to hypoxia and anoxia.[6,55] It is clear that birth complications interact with psychosocial risk factors, like maternal stress, poor parenting and an unstable family environment.[32]

In summary, prenatal development seems to affect brain development and as such affects behaviour, which in some cases results in violent behaviour later in life. Minor physical anomalies may be considered as markers of deviant brain development during pregnancy.[6] Features like low-set ears, adherent ear lobes and a furrowed tongue are anomalies that have been described [6] and shown to predict violent offending in unstable home situations.[6,32] A schematic overview of these factors is given in table 5.

Table 5. Overview of the evidence for involved prenatal environmental factors.

Prenatal environmental factors	**Population**	**Method**	**Outcomes**	**Reference**
Prenatal alcohol exposure	Pregnant women	Interview, tests	Increased risk for conduct disorder	[6]
Nicotine	Pregnant women	Interview, arrest history	Dose-response relationship between use during pregnancy and violence	[6,55]
Nutrition	Pregnant women	Follow-up of offspring	Nutritional deficits during the first two trimesters had children with antisocial personality disorder more often	[32,66]
Birth complication	Pregnant women	Follow-up of offspring	Anoxia, preeclampsia and forceps delivery lead to increased risk for antisocial and criminal behaviour through brain dysfunction	[6,32]

3.2.4.4.3 Postnatal environmental factors

Although the prenatal environment has an effect on brain development, the post-natal environment also shapes brain functioning and gene expression. In the following paragraphs, several examples show how the environment may interact with genes or brain development to influence the development of aggressive or violent criminal behaviour.

3.2.4.4.3.1 Age

An explanation for the robust observation of age as a risk factor for criminal behaviour is found in the development of the prefrontal cortex. [6,32,90,109] A first explanation states that since the myelinisation of the prefrontal cortex continues into a person's 20s or even 30s, it simply cannot cope with the executive demands of adulthood placed upon an individual after adolescence. [6,7,32]

A second explanation is found in the accessibility of the means, opportunity and motive for aggressive behaviour. [47,57] During adolescence, people first experience significant physical strength and cognitive challenges, are less inhibited by supervision and experience pressure to perform both in school and relationships. [6,57,90]

The combination of these hypotheses offers more insight. The changing environment of adolescence requires increased executive functioning, which relies on the prefrontal cortex. [6] Overload of the prefrontal cortex results in impaired development, leading to antisocial behaviour. A stable, supportive environment may offer protection from this harmful overload. [6]

A third explanation is offered by the influence of testosterone. The high-risk periods of adolescence and young adulthood overlap with a testosterone curve in many cultures. [75] So the peak occurrence of violence at these ages could be caused by testosterone. [75] The peaks in sensation seeking, possibly related to testosterone and cortisol, are also seen during these time periods. [90]

The highest risk of violent behaviour is indeed found in persons in their late teens and early twenties. [4,8,32,57,75,90,99,118,119] This holds for both the general population and people who are mentally ill. [4]

3.2.4.4.3.2 Poor child-rearing

The cycle of violence hypothesis states that a history of growing up in a violent context, [73,84] defined as a genetic predisposition, [114,116] a history of witnessing violence or being victimised [26,73,88] leads to

committing violence, possibly by desensitisation, and an acceptance of violence as normal. [73] This would lead to changed psychophysiological parameters like reduced cortisol and decreased autonomic arousal, possibly through an altered development of the limbic system. [92]

The increased cortisol levels of infants separated from their mothers have in fact been shown. [47] This might lead to abnormalities in the hypothalamic–pituitary–adrenal axis, [10,92] leading to hippocampal atrophy, based on stress caused by a lack of affect or traumatic childhood experiences. [92] This in turn has been hypothesised to lead to more proactive aggression. [73]

This theory is supported by various backgrounds that have been found to influence or predict behaviour. For instance child abuse, [3,4,8,17,22,26,31,43,55,57,70,73,92,112,116,120–122] witnessed violence [10,73] or domestic violence, [3,4,31,55,57,111,120,123] family criminality, [4,45,55,99,101,120] marital conflict, [55] early puberty timing, [90,109] early sexual activity, [85] teenage pregnancy, [55] negative emotional attitude from parents [47,73] or mother [57,92,101,112,114,122] and physical maltreatment [73,111,114,120] are all correlated with crime and antisocial behaviour. Also, a child's antisocial behaviour [55] or hyperactivity-impulsivity-attention deficit [55] predicts later criminal behaviour. Some of these correlations exist through direct influence, indirect influence or function as a marker. [85]

As another example, sexual abuse as a child has been associated with later alcohol dependence. [124] This is a risk factor leading to violence. [124] A reduction in child abuse by 50% can be achieved by simple home visits during the first two years of child rearing. [57] Community-based programs also improve self-reported well-being of the parents. [57] Given the influence child abuse has on developing antisocial and offensive behaviour, these programs could lower crime rates.

3.2.4.4.3.3 Socioeconomic status

The lower socioeconomic classes, as measured by SES, are overrepresented among criminals, and a direct correlation has been found. [8,17,70,78,100,119,123] Sub factors of the SES classification, like poverty, [3,17,55] unemployment [52] and school failure [55,99,120,125] have also been found to correlate with criminal behaviour. As an explanation for this finding, the increased stress caused

by low economic status has been mentioned. [78,88] In addition, serotonin response correlates with a SES-score, therefore, a lack of serotonin could confound the correlation between low SES-scores and criminality as well. [78] However, the increased need for and acceptance of violence are also mentioned. In adolescents, it was found that subjects with either high or low social status were more inclined to use physical aggression at school. Middle economic status was a protective factor. [78]

3.2.4.4.3.4 Low IQ

Low IQ-scores, [10,17,57,99,119,120,125] especially for verbal intelligence, [43,50,99] form a risk factor for delinquency and antisocial behaviour. [10] One explanation is the expected lower achievement in school, possibly leading to exclusion, poverty and antisocial behaviour. [57] The increased risk of getting caught if one has a low IQ, or an inherent neurobiological correlate between IQ and delinquent behaviour could also account for this finding.

3.2.4.4.3.5 Gang membership

The association between gang membership and delinquency has been established in multiple studies, for both male and female gangs. [10,99] Neurobehavioural deficits, such as a history of head injury or intermittent explosive disorder are found in gang members more often than in controls.[6] When corrected for other risk factors before and after membership, this association still exists. [10] It may be that individuals with neurobehavioral deficits are more likely to become a gang member (because of traits like sensation seeking) or gang membership affects brain functioning.

3.2.4.4.3.6 Nutritional influences

Postnatal nutritional factors and antisocial or violent behaviour are correlated. [55] For example, protein under-nutrition leads to antisocial personality disorders. [32,55] Serotonin depletion, due to tryptophan under-nutrition (the limiting amino-acid which is used for serotonin) caused aggression under laboratory conditions, compared to well-fed controls. [2] This was also found in rats and monkeys. [55]

Iron deficiency has been found in aggressive children and those with a conduct disorder. [55] In children with attention deficit-hyperactivity disorder (ADHD), both a behavioural and cognitive improvement were found when iron was supplemented. [55] High serum copper levels and high hair levels of manganese, lead and cadmium have

been found in aggressive persons. [55] For some of these metals, this effect was only found in combination with low calcium levels. [55]

Though not well understood, the relationship between these metals and behaviour is thought to lie in neurotransmitters. [55] The influence of metals on behaviour is debated though, since few studies have been conducted, not all results have been replicated and no prospective study or study taking other risk factors into account has been published. [55]

3.2.4.4.3.7 Brain damage

The example of Phineas Gage is often used to illustrate the effects of brain damage. His prefrontal cortex was selectively damaged by an iron spike. [1,3,6,21,24,29,35,37,39,70] Though he survived, his behaviour changed after the accident; he became more aggressive and socially inappropriate. [1,3,21,29] Head injury is found in offenders more often than in the general population, [15,35,51,100] and those with prefrontal damage exhibit aggression more often than those without. [13,44,53]

The exact location of the injury influences the changes in behaviour.[50] Dorsal lesions lead to pseudo-depression, marked by apathy and impaired long-term planning; [37,68] orbital lesions lead to more superficial emotional responses and pseudo-psychopathy.[8,37,68] Whether prefrontal cortex damage leads to criminality or socially less accepted behaviour, is not yet predictable. [1]

The age of injury also has an influence. [50] When the prefrontal cortex injury occurs before adolescence, it leads to diminished executive functioning and what is called 'acquired sociopathy'. [1] When the injury happens in adulthood, however, more impulsive and uncontrolled emotional behaviour results, but executive functioning is not reduced.[1] The age at the time of brain damage also predicts the age for the start of the criminal career of offenders. [66]

However, as mentioned before, there are also case reports of people suffering from the same brain injuries as offenders, who do not show violent aggressive behaviour. [21] Overall, several environmental factors are involved in aggressive behaviour and criminality, some more understood and replicated than others. A schematic overview is given in table 6.

Table 6. Overview of the evidence for environmental factors.

Postnatal environmental factors	Population	Method	Outcomes	Reference
Age	Humans	Database search	Highest risk for violent behaviour is in late teens and early twenties	[4,8,32,57,75,90,99,118,119]
Child abuse	Various	Various	Crime and antisocial behaviour	[3,4,8,17,22,26,31,43,55,57,70,73,92,112,116,120–122]
Antisocial behaviour	Children	Follow-up	Predicts later criminal behaviour	[55]
Hyperactivity-impulsivity-attention deficit	Children	Follow-up	Predicts later criminal behaviour	[55]
Socioeconomic status	Humans	Various	Direct correlation with criminality	[8,17,70,78,100,119,123]
Low IQ-scores	Humans	Various	Risk factor for delinquency and antisocial behaviour	[10,17,57,99,119,120,125]
Gang membership	Humans	Various	Correlation with delinquency	[10,99]
Iron deficiency	Aggressive and conduct disordered children, juvenile delinquents	Plasma levels	Iron deficiency was overrepresented	[32,55]

3.2.5 Discussion

The aim of this paper was to review evidence of biological risk factors that predispose individuals to antisocial and violent behaviour, and to discuss their use for forensic assessment. Several aspects that complicate comparing research in this area must be mentioned to understand the usefulness of the reviewed evidence.

First, much research in this field is performed on psychiatric patients or normal populations, not on offenders. Although the number of studies using groups of offenders grew between 2000 and 2013, there

is still a great need to understand specific offender subgroups. Even if studies use offenders, most groups studied fail to represent the entire imprisoned population. Different studies each select different offender groups thus making the results less valid. [50,53,69,105,120] Defining the studied population is difficult, and different choices are the cause of many differences between studies. In addition, many of the groups studied are simply too small to draw any meaningful conclusions that extend to all offenders, [3,32,50] or to find reliable results that can be replicated.

Second, most studies, specifically neuroimaging studies, compare groups of offenders with other groups of individuals. Forensic (risk) assessments mainly focus on a relationship between deviances and violent behaviour shown by a single individual when committing a crime. Therefore, the forensic field is in need of research showing how alterations in genes, brain, or psychophysiology influence violent behaviour in a specific individual at a specific moment in time.

Third, apart from research on gene-environment interactions, studies on the relationship between neurobiological deficits and violent behaviour that also take psychological or sociological evidence into account are scarce. Most reviewed primary research focuses on only one of the fields related to violent aggressive behaviour, not on the interaction between these fields. Violent aggression, like all forms of human behaviour, [112] does not only develop under specific genetic and environmental conditions, but rather it requires an interplay between the two. [7,69,76,95,101,103,108,110,126] Violence should be considered as the end product of a chain of events over the course of a person's development, during which risks accumulate and potentially reinforce each other. [57] This research gap should be bridged.

Fourth, the interaction does not lead directly to violent aggressive behaviour, but to sensation seeking, impulsivity or low harm avoidance. Evidence of alterations that solely explain violent behaviour was not found. Therefore, it is unlikely that genetic or neuroscientific tools will be used as independent tests in forensic (risk) assessments.

Fifth, studies that do relate neurobiological deficits to behaviour use a variety of aggressive or antisocial behaviours that are not necessarily of use for forensic assessment, which is mainly interested in physical or violent aggression. How violence, aggression and

delinquency are defined and quantified differs in every test; and self-report scales are unreliable. [96,105,120,126] In addition, the distinction between violent reactive and instrumental aggression is not always clear, although these forms of violence are likely to have very disparate neurological backgrounds. [2,10,44,76,101]

Sixth, different studies use a variety of techniques and methods. Neuroanatomical studies focus on imaging single subjects so it is hard to place the subject in a context where violence is likely to be triggered. Neuropsychological studies, on the other hand, often use large populations and are able to test subjects in more ecologically valid situations.

Specifically, in imaging studies, the various regions of the frontal cortex are usually not considered separately. [27,35,44] Also the nuclei of the amygdala are not measured separately. [54] This leads to generalisation, simplification and reduced power, since only some of these regions might actually be linked to deviant behaviour.

Also, testing levels of substances in subjects differs per study. The circadian rhythm of cortisol is not always taken into account. [72,93] Various time periods between samples and circumstances make studies hard to compare.

To conclude, with better designed studies and more standardisation, comparing studies would be easier and it might become possible to link behaviour to underlying mechanisms. [53]

3.2.6 Conclusion

The influence genes and deviances in brain development have on the development of violent aggressive behaviour, and in which situation, needs further research before genetic and brain imaging information can be used in forensic assessments or in court. [11,20,76] Though most mechanisms are not elucidated, some of the findings may in time be used to estimate risk of recidivism in combination with psychosocial assessment tools. This means better tools for neurologically based assessment might become available as the knowledge develops.

As the developmental profile of brain areas and their vulnerabilities are being discovered, key moments to modulate specific environmental factors for persons with a high-risk genetic profile will become possible. [114] For example, some findings can be used to more

accurately assess risk of criminal behaviour on an individual basis. However, there is an important ethical difference between using neurobiological assessment tools in the case of suspects and convicted offenders versus in the general population or subgroups, such as children or adolescents. Even in case of the former group, offender rights might be at stake. [65]

On a more general level, knowledge of nutrition could be used to improve our society or correctional facilities, and help prevent future encounters with forensic facilities. Better guidance during the most difficult years of adolescence and home visits can diminish chances of a harmful overload of the prefrontal cortex and decrease chances of child abuse. And obviously, brain damage should be avoided. Reducing those criminogenic risk factors reduces the likelihood of engaging in criminal activity, both directly and via reduced triggering of gene-environment interactions. [103] In the future, new information from neuroscience, when integrated into the information already available from sociological and psychological assessments, could contribute to the development of better risk assessment tools, treatments and cures for offenders, reducing recidivism as well. [16,21,63,66]

This review underlines the importance of maintaining a case-by-case differentiated approach to evidence-based forensic assessment that takes into account the individual psychosocial development, and neurobiological and genetic risk factors contributing to violent crime.

3.2.7 Acknowledgements

The authors would like to thank Tale Evenhuis, information specialist for the extensive technical support during the systematic review process and Julia Challinor as well as Nathalie Kuijpers for their English manuscript correction services. We are also most grateful to the three anonymous reviewers for their invaluable comments.

3.2.8 References

1. Sapolsky RM (2004) The frontal cortex and the criminal justice system. Philosophical Transactions of the Royal Society B: Biological Sciences 359: 1787-1796.
2. Davidson RJ, Putnam KM, Larson CL (2000) Dysfunction in the neural circuitry of emotion regulation-a possible prelude to violence. Science 289: 591-594.
3. Hoptman MJ (2003) Neuroimaging studies of violence and antisocial behavior. Journal of Psychiatric Practice 9: 265-278.

4. Anderson TR, Bell CC, Powell TE, Williamson JL, Blount J (2004) Assessing psychiatric patients for violence. Community Mental Health Journal 40: 379-399.
5. Pallone NJ, Hennessy JJ (2000) Neuropathology and criminal violence: Newly calibrated ratios. Journal of Offender Rehabilitation 31: 87-99.
6. Raine A (2002) Biosocial studies of antisocial and violent behavior in children and adults: a review. Journal of Abnormal Child Psychology 30: 311-326.
7. Rothstein MA (2005) Science and society: applications of behavioural genetics: outpacing the science? Nature reviews Genetics. 6: 793-798.
8. Denney RL, Wynkoop TF (2000) Clinical neuropsychology in the criminal forensic setting. Journal of Head Trauma Rehabilitation 15: 804-828.
9. Heilbronner RL (2004) A status report on the practice of forensic neuropsychology. Clinical Neuropsychologist 18: 312-326.
10. Rappaport N, Thomas C (2004) Recent research findings on aggressive and violent behavior in youth: Implications for clinical assessment and intervention. Journal of Adolescent Health 35: 260-277.
11. Forzano F, Borry P, Cambon-Thomsen A, Hodgson SV, Tibben A, de VP, van EC, Cornel M (2010) Italian appeal court: a genetic predisposition to commit murder? Eur J Hum Genet 18: 519-521. http://dx.doi.org/10.1038/ejhg.2010.31.
12. Morse SJ (2011) Genetics and criminal responsibility. Trends in Cognitive Sciences 15: 378-380.
13. Penney S (2012) Impulse control and criminal responsibility: Lessons from neuroscience. International Journal of Law and Psychiatry 35: 99-103. http://dx.doi.org/10.1016/j.ijlp.2011.12.004.
14. Beszterczey S, Nestor PG, Shirai A, Harding S (2013) Neuropsychology of decision making and psychopathy in high-risk ex-offenders. Neuropsychology 27: 491-497. 2013-25138-009 [pii];10.1037/a0033162 [doi].
15. Casartelli L, Chiamulera C (2013) Opportunities, threats and limitations of neuroscience data in forensic psychiatric evaluation. Current Opinion in Psychiatry 26: 468-473.
16. May JS, Beaver KM (2014) The neuropsychological contributors to psychopathic personality traits in adolescence. Int J Offender Ther Comp Criminol 58: 265-285. 0306624X12469861 [pii];10.1177/0306624X12469861 [doi].
17. Miller HV, Barnes JC (2013) Genetic transmission effects and intergenerational contact with the criminal justice system: A consideration of three dopamine polymorphisms. Criminal Justice and Behavior Vol.40: 671-689.

18. Berryessa CM, Cho MK (2013) Ethical, legal, social, and policy implications of behavioral genetics. Annual Review of Genomics and Human Genetics 14: 515-534.
19. Hughes V (2010) Science in court: head case. Nature 464: 340-342. http://dx.doi.org/10.1038/464340a.
20. Baum ML (2013) The monoamine oxidase A (MAOA) genetic predisposition to impulsive violence: Is it relevant to criminal trials? Neuroethics 6: 287-306.
21. Schleim S (2012) Brains in context in the neurolaw debate: The examples of free will and "dangerous" brains. International Journal of Law and Psychiatry 35: 104-111.
22. Ward T, Beech A (2006) An integrated theory of sexual offending. Aggression and Violent Behavior 11: 44-63.
23. Anckarsater H (2006) Central nervous changes in social dysfunction: Autism, aggression, and psychopathy. Brain Research Bulletin 69: 259-265.
24. Goodenough OR, Prehn K (2004) A neuroscientific approach to normative judgment in law and justice. Philosophical Transactions of the Royal Society B: Biological Sciences 359: 1709-1726.
25. Glenn AL, Raine A (2008) The Neurobiology of Psychopathy. Psychiatric Clinics of North America 31: 463-475.
26. Mitchell IJ, Beech AR (2011) Towards a neurobiological model of offending. Clinical Psychology Review 31: 872-882.
27. Golkar A, Lonsdorf TB, Olsson A, Lindstrom KM, Berrebi J, Fransson P, Schalling M, Ingvar M, +ûhman A (2012) Distinct Contributions of the Dorsolateral Prefrontal and Orbitofrontal Cortex during Emotion Regulation. PLoS ONE 7: e48107. doi:10.1371/journal.pone.0048107.
28. Wahlund K, Kristiansson M (2009) Aggression, psychopathy and brain imaging - Review and future recommendations. Int J Law Psychiatry 32: 266-271. http://dx.doi.org/10.1016/j.ijlp.2009.04.007.
29. Marazziti D, Baroni S, Landi P, Ceresoli D, Dell'Osso L (2013) The neurobiology of moral sense: Facts or hypotheses? Annals of General Psychiatry 12.
30. Victor TA, Furey ML, Fromm SJ, Bellgowan PSF, +ûhman A, Drevets WC (2012) The Extended Functional Neuroanatomy of Emotional Processing Biases for Masked Faces in Major Depressive Disorder. PLoS ONE 7: e46439. doi:10.1371/journal.pone.0046439.
31. Dutton DG (2002) The neurobiology of abandonment homicide. Aggression and Violent Behavior 7: 407-421.
32. Raine A (2002) Annotation: The role of prefrontal deficits, low autonomic arousal and early health factors in the development of antisocial and aggressive behavior in children. Journal of Child Psychology and Psychiatry 43: 417-434.

33. Beckham JC, Moore SD, Reynolds V (2000) Interpersonal hostility and violence in Vietnam combat veterans with chronic posttraumatic stress disorder: A review of theoretical models and empirical evidence. Aggression and Violent Behavior 5: 451-466.
34. Siegel A, Douard J (2011) Who's flying the plane: serotonin levels, aggression and free will. Int J Law Psychiatry 34: 20-29. http://dx.doi.org/10.1016/j.ijlp.2010.11.004.
35. Schiltz K, Witzel JG, Bausch-Holterhoff J, Bogerts B (2013) High prevalence of brain pathology in violent prisoners: A qualitative CT and MRI scan study. European Archives of Psychiatry and Clinical Neuroscience Vol.263: 607-616.
36. Takahashi A, Nagayasu K, Nishitani N, Kaneko S, Koide T (2014) Control of Intermale Aggression by Medial Prefrontal Cortex Activation in the Mouse. PLoS ONE 9: e94657. doi:10.1371/journal.pone.0094657.
37. Stein DJ (2000) The neurobiology of evil: Psychiatric perspectives on perpetrators. Ethnicity & Health 5: 303-315.
38. Pontius AA (2005) Fastest fight/flight reaction via amygdalar visual pathway implicates simple face drawing as its marker: Neuroscientific data consistent with neuropsychological findings. Aggression and Violent Behavior 10: 363-373.
39. Gao Y, Glenn AL, Schug RA, Yang Y, Raine A (2009) The neurobiology of psychopathy: a neurodevelopmental perspective. Can J Psychiatry 54: 813-823.
40. Viding E (2004) Annotation: Understanding the development of psychopathy. Journal of Child Psychology and Psychiatry and Allied Disciplines 45: 1329-1337.
41. Martens WHJ (2002) Criminality and moral dysfunctions: Neurological, biochemical, and genetic dimensions. International Journal of Offender Therapy and Comparative Criminology 46: 170-182.
42. Laurell J, Daderman AM (2005) Recidivism is related to psychopathy (PCL-R) in a group of men convicted of homicide. International Journal of Law and Psychiatry 28: 255-268.
43. Hill J (2002) Biological, psychological and social processes in the conduct disorders. Journal of Child Psychology and Psychiatry and Allied Disciplines 43: 133-164.
44. Blair RJR (2005) Applying a cognitive neuroscience perspective to the disorder of psychopathy. Development and Psychopathology 17: 865-891.
45. McEllistrem JE (2004) Affective and predatory violence: A bimodal classification system of human aggression and violence. Aggression and Violent Behavior 10: 1-30.
46. Blackwood NJ (2012) The antisocial brain: Psychopathy matters. Biological Psychiatry Conference: 127S.

47. Meier NM, Perrig W, Koenig T (2012) Neurophysiological correlates of delinquent behaviour in adult subjects with ADHD. International Journal of Psychophysiology 84: 1-16.
48. Nelson RJ, Trainor BC (2007) Neural mechanisms of aggression. Nat Rev Neurosci 8: 536-546. http://dx.doi.org/10.1038/nrn2174.
49. Yang Y, Raine A (2009) Prefrontal structural and functional brain imaging findings in antisocial, violent, and psychopathic individuals: A meta-analysis. Psychiatry Research - Neuroimaging 174: 81-88.
50. Teichner G, Golden CJ (2000) The relationship of neuropsychological impairment to conduct disorder in adolescence: A conceptual review. Aggression and Violent Behavior 5: 509-528.
51. Cohen RA, Brumm V, Zawacki TM, Paul R, Sweet L, Rosenbaum A (2003) Impulsivity and verbal deficits associated with domestic violence. Journal of the International Neuropsychological Society 9: 760-770.
52. Wilson LC, Scarpa A (2012) Criminal behavior: The need for an integrative approach that incorporates biological influences. Journal of Contemporary Criminal Justice 28: 366-381. http://dx.doi.org/10.1177/1043986212450232.
53. Brower MC, Price BH (2001) Neuropsychiatry of frontal lobe dysfunction in violent and criminal behaviour: A critical review. Journal of Neurology, Neurosurgery & Psychiatry 71: 720-726.
54. Sugranyes G, Kyriakopoulos M, Corrigall R, Taylor E, Frangou S (2011) Autism spectrum disorders and schizophrenia: meta-analysis of the neural correlates of social cognition. PLoS One 6: e25322. 10.1371/journal.pone.0025322 [doi];PONE-D-11-10972 [pii].
55. Liu J, Wuerker A (2005) Biosocial bases of aggressive and violent behavior--Implications for nursing studies. International Journal of Nursing Studies 42: 229-241.
56. Pridmore S, Chambers A, McArthur M (2005) Neuroimaging in psychopathy. Australian and New Zealand Journal of Psychiatry 39: 856-865.
57. Fonagy P (2004) Early-Life Trauma and the Psychogenesis and Prevention of Violence. Annals of the New York Academy of Sciences 1036: 181-200.
58. Reidy DE, Shelley-Tremblay JF, Lilienfeld SO (2011) Psychopathy, reactive aggression, and precarious proclamations: A review of behavioral, cognitive, and biological research. Aggression and Violent Behavior 16: 512-524.
59. DeLisi M (2009) Psychopathy is the unified theory of crime. Youth Violence and Juvenile Justice 7: 256-273.
60. Dolan M (2012) The neuropsychology of prefrontal function in antisocial personality disordered offenders with varying degrees of psychopathy. Psychological Medicine 42: 1715-1725. http://dx.doi.org/10.1017/S0033291711002686.

61. Sartori G, Pellegrini S, Mechelli A (2011) Forensic neurosciences: From basic research to applications and pitfalls. Current Opinion in Neurology 24: 371-377.
62. Muller JL (2010) Psychopathy--an approach to neuroscientific research in forensic psychiatry. Behavioral Sciences & the Law 28: 129-147.
63. Joyal CC, Plante-Beaulieu J, de CA (2013) The Neuropsychology of Sexual Offenders: A Meta-Analysis. Sex Abuse . 1079063213482842 [pii];10.1177/1079063213482842 [doi].
64. Jordan K, Fromberger P, Stolpmann G, Muller JL (2011) The role of testosterone in sexuality and paraphilia-a neurobiological approach. Part II: Testosterone and paraphilia. Journal of Sexual Medicine 8: 3008-3029.
65. Aharoni E, Vincent GM, Harenski CL, Calhoun VD, Sinnott-Armstrong W, Gazzaniga MS, Kiehl KA (2013) Neuroprediction of future rearrest. Proc Natl Acad Sci U S A 110: 6223-6228. 1219302110 [pii];10.1073/pnas.1219302110 [doi].
66. Glenn AL, Raine A (2014) Neurocriminology: Implications for the punishment, prediction and prevention of criminal behaviour. Nature Reviews Neuroscience 15: 54-63.
67. Baumbach J (2002) Some implications of prenatal alcohol exposure for the treatment of adolescents with sexual offending behaviors. Sexual abuse : a journal of research and treatment 14: 313-327.
68. Morgan AB, Lilienfeld SO (2000) A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. Clinical Psychology Review 20: 113-156.
69. Philipp-Wiegmann F, Romer KD, Retz W, Rosler M (2011) Reduced cortical inhibition in violent offenders: A study with transcranial magnetic stimulation. European Psychiatry Conference.
70. Stone MH (2001) Serial sexual homicide: Biological, psychological, and sociological aspects. Journal of Personality Disorders 15: 1-18.
71. Raine A, Venables PH, Mednick SA (1997) Low resting heart rate at age 3 years predisposes to aggression at age 11 years: evidence from the Mauritius Child Health Project. J Am Acad Child Adolesc Psychiatry 36: 1457-1464. http://dx.doi.org/10.1097/00004583-199710000-00029.
72. Lipschitz DS, Morgan III CA, Southwick SM (2002) Neurobiological disturbances in youth with childhood trauma and in youth with conduct disorder. Journal of Aggression, Maltreatment and Trauma 6: 149-174.
73. Scarpa A (2003) Community violence exposure in young adults. Trauma, Violence, & Abuse 4: 210-227.
74. Nelson RJ, Chiavegatto S (2001) Molecular basis of aggression. Trends in Neurosciences 24: 713-719.
75. Mong JA, Pfaff DW, Sultan, Slob, Swaab, De K (2003) Hormonal and genetic influences underlying arousal as it drives sex and aggression in animal and human brains. Neurobiology of Aging 24: S83-S92.

76. Vassos E, Collier DA, Fazel S (2013) Systematic meta-analyses and field synopsis of genetic association studies of violence and aggression. Mol Psychiatry . 10.1038/mp.2013.31.
77. Koh KB, Choi EH, Lee Y-J, Han M, Choi S-S, Kim SW, Lee MG (2012) The relation of serotonin-related gene and COMT gene polymorphisms with criminal behavior in schizophrenic disorder. Journal of Clinical Psychiatry 73: 159-163.
78. Aslund C, Comasco E, Nordquist N, Leppert J, Oreland L, Nilsson KW (2013) Self-Reported Family Socioeconomic Status, the 5-HTTLPR Genotype, and Delinquent Behavior in a Community-Based Adolescent Population. Aggressive Behavior 39: 52-63.
79. Popova NK (2008) From gene to aggressive behavior: the role of brain serotonin. Neurosci Behav Physiol 38: 471-475. http://dx.doi.org/10.1007/s11055-008-9004-7.
80. Olivier B, van OR (2005) 5-HT1B receptors and aggression: a review. Eur J Pharmacol 526: 207-217. http://dx.doi.org/10.1016/j.ejphar.2005.09.066.
81. Bouvy PF, Liem M (2012) Antidepressants and lethal violence in the Netherlands 1994-2008. Psychopharmacology (Berl) 222: 499-506. http://dx.doi.org/10.1007/s00213-012-2668-2.
82. Beaver KM, Chaviano N (2011) The association between genetic risk and contact with the criminal justice system in a sample of Hispanics. Journal of Contemporary Criminal Justice 27: 81-94. http://dx.doi.org/10.1177/1043986210396205.
83. Levitt M (2013) Genes, environment and responsibility for violent behavior: "Whatever genes one has it is preferable that you are prevented from going around stabbing people". New Genetics and Society 32: 4-17.
84. Janssen PA, Nicholls TL, Kumar RA, Stefanakis H, Spidel AL, Simpson EM (2005) Of mice and men: Will the intersection of social science and genetics create new approaches for intimate partner violence? Journal of Interpersonal Violence 20: 61-71.
85. Harden KP, Mendle J (2011) Adolescent sexual activity and the development of delinquent behavior: The role of relationship context. Journal of Youth and Adolescence 40: 825-838. http://dx.doi.org/10.1007/s10964-010-9601-y.
86. Barnes JC, Jacobs BA (2013) Genetic Risk for Violent Behavior and Environmental Exposure to Disadvantage and Violent Crime: The Case for Gene-Environment Interaction. Journal of Interpersonal Violence 28: 92-120.
87. Dmitrieva J, Chen C, Greenberger E, Ogunseitan O, Ding YC (2011) Gender-specific expression of the DRD4 gene on adolescent delinquency, anger and thrill seeking. Social Cognitive and Affective Neuroscience 6: 82-89. http://dx.doi.org/10.1093/scan/nsq020.

88. Vaske J, Wright JP, Beaver KM (2011) A dopamine gene (DRD2) distinguishes between offenders who have and have not been violently victimized. International Journal of Offender Therapy and Comparative Criminology 55: 251-267. http://dx.doi.org/10.1177/0306624X10361583.
89. Hirata Y, Zai CC, Nowrouzi B, Beitchman JH, Kennedy JL (2013) Study of the Catechol-O-Methyltransferase (COMT) Gene with High Aggression in Children. Aggressive Behavior 39: 45-51.
90. Harden K, Quinn P, Tucker-Drob E (2011) Genetically influenced changes in sensation seeking drive the rise of delinquent behavior during adolescence. Behavior Genetics Conference: 911.
91. Miczek KA, Fish EW, Almeida RMM, Faccidomo S, Debold JF (2004) Role of Alcohol Consumption in Escalation to Violence. Annals of the New York Academy of Sciences 1036: 278-289.
92. Beech AR, Mitchell IJ (2005) A neurobiological perspective on attachment problems in sexual offenders and the role of selective serotonin re-uptake inhibitors in the treatment of such problems. Clinical Psychology Review 25: 153-182.
93. Ramirez JM (2003) Hormones and aggression in childhood and adolescence. Aggression and Violent Behavior 8: 621-644.
94. Spironelli C, Gradante F, Gradante G, Angrilli A (2013) Cortisol and ACTH plasma levels in maternal filicides and violent psychiatric women. J Psychiatr Res 47: 622-627. http://dx.doi.org/10.1016/j.jpsychires.2013.01.001.
95. Aluja A, Luis F, Blanch A, Fibla J (2011) Association of androgen receptor gene, CAG and GGN repeat length polymorphism and impulsive-disinhibited personality traits in inmates: the role of short-long haplotype. Psychiatric Genetics 21: 229-239. http://dx.doi.org/10.1097/YPG.0b013e328345465e.
96. Tikkanen R, Auvinen-Lintunen L, Ducci F, Sjoberg RL, Goldman D, Tiihonen J, Ojansuu I, Virkkunen M (2011) Psychopathy, PCL-R, and MAOA genotype as predictors of violent reconvictions. Psychiatry Research 185: 382-386.
97. Berman ME (2000) Accruing knowledge on the biological bases of aggression: Comment on Giancola (2000). Experimental and Clinical Psychopharmacology 8: 601-603.
98. Bourget D, Whitehurst L (2004) Capgras Syndrome: A Review of the Neurophysiological Correlates and Presenting Clinical Features in Cases Involving Physical Violence. The Canadian Journal of Psychiatry / La Revue canadienne de psychiatrie 49: 719-725.
99. Kurtz A (2002) What works for delinquency? The effectiveness of interventions for teenage offending behaviour. Journal of Forensic Psychiatry 13: 671-692.

100. Hanlon RE, Coda JJ, Cobia D, Rubin LH (2012) Psychotic domestic murder: Neuropsychological differences between homicidal and nonhomicidal schizophrenic men. Journal of Family Violence 27: 105-113. http://dx.doi.org/10.1007/s10896-011-9410-4.
101. DiLalla LF (2002) Behavior genetics of aggression in children: Review and future directions. Developmental Review 22: 593-622.
102. Arias JMC, Acosta CAP, Valencia JG, Montoya GJ, Viana JCA, Nieto OC, Florez AF, Medellin BEC, Montoya WR, Jaramilloa CAL, Achury JG, Fuentes CC, Berrio GB, Ruiz-Linares A (2011) Exploring epistasis in candidate genes for antisocial personality disorder. Psychiatric Genetics 21: 115-124. http://dx.doi.org/10.1097/YPG.0b013e3283437175.
103. Beaver KM (2011) Environmental moderators of genetic influences on adolescent delinquent involvement and victimization. Journal of Adolescent Research 26: 84-114. http://dx.doi.org/10.1177/0743558410384736.
104. Schwartz JA, Beaver KM (2013) Exploring whether genetic differences between siblings explain sibling differences in criminal justice outcomes. Compr Psychiatry . S0010-440X(13)00139-9 [pii];10.1016/j.comppsych.2013.06.002 [doi].
105. Kendler KS, Patrick CJ, Larsson H, Gardner CO, Lichtenstein P (2013) Genetic and environmental risk factors in males for self-report externalizing traits in mid-adolescence and criminal behavior through young adulthood. Psychol Med 43: 2161-2168. S003329171300007X [pii];10.1017/S003329171300007X [doi].
106. Beaver KM, Wright JP, Boutwell BB, Barnes JC, DeLisi M, Vaughn MG (2013) Exploring the association between the 2-repeat allele of the MAOA gene promoter polymorphism and psychopathic personality traits, arrests, incarceration, and lifetime antisocial behavior. Personality and Individual Differences 54: 164-168. http://dx.doi.org/10.1016/j.paid.2012.08.014.
107. Boutwell BB, Menard S, Barnes JC, Beaver KM, Armstrong TA, Boisvert D (2013) The role of gene-gene interaction in the prediction of criminal behavior. Compr Psychiatry . S0010-440X(13)00329-5 [pii];10.1016/j.comppsych.2013.11.005 [doi].
108. Vaske J, Newsome J, Wright JP (2012) Interaction of serotonin transporter linked polymorphic region and childhood neglect on criminal behavior and substance use for males and females. Development and Psychopathology 24: 181-193. http://dx.doi.org/10.1017/S0954579411000769.
109. Harden KP, Mendle J (2012) Gene-environment interplay in the association between pubertal timing and delinquency in adolescent girls. Journal of Abnormal Psychology 121: 73-87. http://dx.doi.org/10.1037/a0024160.

110. Chen B, Gu T, Ma B, Zheng G, Ke B, Zhang X, Zhang L, Wang Y, Hu L, Chen Y, Qiu J, Nie S (2013) The CRHR1 Gene Contributes to Genetic Susceptibility of Aggressive Behavior Towards Others in Chinese Southwest Han Population. J Mol Neurosci . 10.1007/s12031-013-0160-z [doi].
111. Rhee SH, Waldman ID (2002) Genetic and environmental influences on antisocial behavior: a meta-analysis of twin and adoption studies. Psychological Bulletin 128: 490-529.
112. Schwartz JA, Beaver KM (2011) Evidence of a gene x environment interaction between perceived prejudice and MAOA genotype in the prediction of criminal arrests. Journal of Criminal Justice 39: 378-384. http://dx.doi.org/10.1016/j.jcrimjus.2011.05.003.
113. Vaske J, Wright JP, Boisvert D, Beaver KM (2011) Gender, genetic risk, and criminal behavior. Psychiatry Research 185: 376-381.
114. Viding E (2004) On the Nature and Nurture of Antisocial Behavior and Violence. Annals of the New York Academy of Sciences 1036: 267-277.
115. Connolly EJ, Beaver KM (2014) Examining the genetic and environmental influences on self-control and delinquency: results from a genetically informative analysis of sibling pairs. J Interpers Violence 29: 707-735. 0886260513505209 [pii];10.1177/0886260513505209 [doi].
116. Wasserman D (2004) Is there value in identifying individual genetic predispositions to violence? The Journal of law, medicine & ethics : a journal of the American Society of Law, Medicine & Ethics 32: 24-33.
117. Hall J, Philip RCM, Marwick K, Whalley HC, Romaniuk L, McIntosh AM, Santos I, Sprengelmeyer R, Johnstone EC, Stanfield AC, Young AW, Lawrie SM (2012) Social Cognition, the Male Brain and the Autism Spectrum. PLoS ONE 7: e49033. doi:10.1371/journal.pone.0049033.
118. Pope K, Luna B, Thomas CR (2012) Developmental neuroscience and the courts: How science is influencing the disposition of juvenile offenders. Journal of the American Academy of Child & Adolescent Psychiatry 51: 341-342. http://dx.doi.org/10.1016/j.jaac.2012.01.003.
119. Barnes JC (2013) Analyzing the origins of life-course-persistent offending: A consideration of environmental and genetic influences. Criminal Justice and Behavior Vol.40: 519-540.
120. Brewer-Smyth K (2004) Women behind bars: Could neurobiological correlates of past physical and sexual abuse contribute to criminal behavior? Health Care for Women International 25: 835-852.
121. Silva JA, Leong GB, Ferrari MM (2004) A neuropsychiatric developmental model of serial homicidal behavior. Behavioral Sciences and the Law 22: 787-799.
122. Kunst JL (2002) Fraught with the utmost danger: The object relations of mothers who kill their children. Bulletin of the Menninger Clinic 66: 19-38.

123. Hines DA, Saudino KJ (2002) Intergenerational transmission of intimate partner violence: A behavioral genetic perspective. Trauma, Violence, & Abuse 3: 210-225.
124. Copeland WE, Magnusson A, Goransson M, Heilig MA (2011) Genetic moderators and psychiatric mediators of the link between sexual abuse and alcohol dependence. Drug Alcohol Depend 115: 183-189. S0376-8716(10)00400-X [pii];10.1016/j.drugalcdep.2010.10.024 [doi].
125. Ferguson CJ, Ivory JD, Beaver KM (2013) Genetic, maternal, school, intelligence, and media use predictors of adult criminality: A longitudinal test of the catalyst model in adolescence through early adulthood. Journal of Aggression, Maltreatment & Trauma Vol.22: 447-460.
126. Boisvert D, Vaske J, Wright JP, Knopik V (2012) Sex differences in criminal behavior: A genetic analysis. Journal of Contemporary Criminal Justice 28: 293-313. http://dx.doi.org/10.1177/1043986212450224.

3.3 Evaluation

3.3.1 Strengths

The main strength of this review is the purposely selected search methodology to make the review of a large body of literature possible. The inclusion of literature from multiple scientific disciplines allowed for the identification of overarching themes and links between them.

The double methodology, first searching for other reviews and then adding more recent primary research that would not have been captured, allowed for a balance between being comprehensive and up-to-date, one of the main difficulties in SRs. This does require us to initially accept the selection and assessments of other authors for research that happened before our primary search.

Duplicating the selection, by having two independent reviewers select and discuss the differences, ensured the selection was considered properly and appropriate for the research question. Though it required long conversations about why we had selected specific records, it yielded a robust selection of articles to be included.

Including all three search engines did help with capturing all the relevant information. Out of those 3508 records, only 299 were duplicates. This highlights that the use of each of these databases improved our recall and added value to our SR.

3.3.2 Limitations

Using search terminology aligned with thesauri probably helped in optimizing the search strategy in terms of recall (discussed in paragraph 2.2.3.1). It did not appear to optimize precision, as it produced 3508 hits out of which we selected 147 abstracts. We consider if it might have been easier to use a broader search strategy, based on more high-level terms like “criminology,” “aggression” and “violence,” and select from that based on titles and abstracts.

We can retrospectively search with this approach, and compare what results we would have found. So using this simplified search, *(criminality OR aggression OR antisocial behaviour OR psychopathy) AND (neurosciences OR genetics)*, limited to reviews and the time frame 2000-2013, PsycInfo produces 1125 hits, Embase 1338 and PubMed 4325. The second search for primary research, with the same terms but not limited to reviews and searching only in 2010-2013, produced 1217 records in

PsycInfo, 3076 in Embase and 9716 in PubMed. There are likely many duplicates in these results, but it is safe to say that these 20797 hits in total are significantly more than the 3508 we used now.

We could try to tailor this simplified search further with other limits, fewer years or search terms, but this would have reduced our sensitivity. It might be possible to improve on this search to compile one that yields a similar set of articles without using the thesauri, but this comes at significant risk of missing key papers or vastly increasing the number of records that need to be searched. In hindsight, this search methodology was probably the best one possible.

We did not systematically judge the quality of each study. We discussed likely methodological causes where studies reported conflicting outcomes, but we did not consistently consider the quality of the methodology of each study. It was also not possible for us to systematically assess publication bias, as the publications we identified were not comparable enough to analyse through a funnel plot. This means that our results might be influenced by biases we are not aware of.

As this SR used data from several disciplines, there were notable differences in the patient populations and definitions used to describe them. Some trials worked with healthy participants, others only with ones with a specific diagnosis, and others only with offenders. Timing of samples that were taken from participants varied, making them less comparable for those compounds where diurnal variation is relevant. Sample sizes, specifically for imaging studies, were small, but often extrapolated to much larger populations. Known other risk factors were not consistently taken into account. Methodological shortcomings of the included trials limit the interpretation of the synthesis. A meaningful MA or detection of bias was therefore not possible, and the conclusions are cautious. As with any review, applying the findings to individual cases requires careful consideration of the appropriateness.

3.3.3 Appropriateness of methodology

The objective of this review was to provide an overview of the current evidence about biological risk factors that predispose people to antisocial and violent behaviour, and determine its usefulness in forensic assessment. Given the broad question, spanning many scientific disciplines, our broad approach was appropriate. The fine-grained

selection of each individual term in the thesauri of these search engines took significant effort, but yielded much more specific results than a simplified search string would have produced. Using several search engines and independent, duplicated selection helped with a robust identification and selection of material. The duplicate search, with first only reviews from 2000 to 2013 and then primary research from 2010 to 2013, made sure we balanced capturing the breadth of the research landscape and the most recent developments.

We did not assess study-level biases or publication bias, which would have been a good addition to more reliably draw conclusions about the likely impact of each risk factor.

3.3.4 Lessons learned

If we were to perform this SR again, we would again use a thesaurus-based approach. We might consider reducing the number of years we search for or limit it differently, but selecting each term itself was worth the invested time and effort.

We could also focus on primary research only, to be less reliant on the methodology of previous reviewers. This would have required significantly more time but would have allowed for a more reliable overview, as we could judge individual papers, and would possibly allow for a less cautious conclusion. We would then also assess each study for methodological quality.

This review was viewed almost 10,000 times up to August 2019 according to PLoS One, showing the high visibility of the findings. Google scholar indicated 23 citations in the same period.

This research shows that criminality is the result of a chain of risk factors, and that not one scientific discipline can explain it completely. This helps with forensic risk assessments, but also with the design of future research. Observational research should strive to take all relevant risk factors into account to maximise the chances of finding reliable and reproducible results. Future SRs, possibly with meta-analyses, could focus on aspects of this review, such as only genetics or imaging. A meta-regression analysis might help in identifying which of the risk factors contribute most to the overall risk of developing antisocial and violent behaviour.

In conclusion, this method worked well for this research question. Selecting thesaurus terms at a granular level is a feasible approach, and is applicable to other research questions as well, but requires significant resources and its benefit will depend on the research area. In the next chapter, we will look at another aspect of the search methodology.

4 Addressing the challenge of high-priced prescription drugs in the era of precision medicine: A systematic review of drug life cycles, therapeutic drug markets and regulatory frameworks

Similar to the previous chapter, this chapter addresses the search and selection methodology of a tailor-made SR. However, it does not consider the search terms, but focuses on another fundamental design choice: the search engines and sources that should be used.

As described in paragraph 2.2.3.1, there are many search engines for SRs, all with different benefits. So, to make sure that the right type of information is found, it is necessary to consider which database is most likely to contain the information that is needed and design a search strategy for each of the chosen search engines.

We investigated a topic that is covered both by the scientific literature and by public media. To fully capture the debate, a mixed-method search in both domains was necessary, so the search strategy was tailored to this. We searched in both scientific journals and selected high-quality newspapers. This allowed for a balanced perspective on the debate.

4.1 Context

This SR discusses a politically sensitive topic: the recent outcry about drug pricing. The aim of the review was to determine what has caused the recent rise in drug prices, and what can be done in terms of policy around drug pricing to safeguard equitable access to innovative medicines.

To answer this, a search in PubMed was performed. Due to the many papers that focus on cost-effectiveness of one or a handful of treatments, those papers were excluded in the search. The final search string was *(Drugs OR medicines) AND (prices OR costs) NOT (Cost-effectiveness OR Clinical Trial OR treatment)*, with limits set to papers published between January 1, 2014 and January 1, 2017 (3 years). After selection and exclusion, this yielded 104 scientific articles.

One downside of using scientific, peer-reviewed papers, is that the delay in publication means that the information can be outdated by

the time it is published. As such, using only scientific papers would mean that this paper gives an overview with a delay, so recent developments would not be discussed. Furthermore, scientific publications might not capture the emotions and patient perspective of this debate.

Therefore, to make sure a comprehensive and up-to-date picture could be painted, newspapers were included in the search. We used both American and British newspapers, to provide input from both sides of the Atlantic, and both business-focussed and general newspapers. The final selection was the Financial Times (FT), the New York Times (NYT), the Guardian and the Wall Street Journal (WSJ). The timeframe was limited to the period between January 1, 2015 and January 1, 2017 (2 years).

4.2 Full text

A preliminary version of this chapter was published in PLoS One.

van der Gronde, T., Uyl-de Groot, C. A., & Pieters, T. (2017). Addressing the challenge of high-priced prescription drugs in the era of precision medicine: A systematic review of drug life cycles, therapeutic drug markets and regulatory frameworks. *PloS One*, 12(8), e0182613. https://doi.org/10.1371/journal.pone.0182613

Addressing the challenge of high-priced prescription drugs in the era of precision medicine: A systematic review of drug life cycles, therapeutic drug markets and regulatory frameworks

Toon van der Gronde[1], Carin A. Uyl-de Groot[2], Toine Pieters[1]*

[1] Department of Pharmaceutical Sciences, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Utrecht, the Netherlands,

[2] Institute for Medical Technology Assessment, Department of Health Policy & Management, Erasmus University, Rotterdam, the Netherlands

* t.pieters@uu.nl

4.2.1 Abstract

Context Recent public outcry has highlighted the rising cost of prescription drugs worldwide, which in several disease areas outpaces other health care expenditures and results in a suboptimal global availability of essential medicines.

Method A systematic review of Pubmed, the Financial Times, the New York Times, the Wall Street Journal and the Guardian was performed to identify articles related to the pricing of medicines.

Findings Changes in drug life cycles have dramatically affected patent medicine markets, which have long been considered a self-evident and self-sustainable source of income for highly profitable drug companies. Market failure in combination with high merger and acquisition activity in the sector have allowed price increases for even off-patent drugs. With market interventions and the introduction of QALY measures in health care, governments have tried to influence drug prices, but often encounter unintended consequences. Patent reform legislation, reference pricing, outcome-based pricing and incentivizing physicians and pharmacists to prescribe low-cost drugs are among the most promising short-term policy options. Due to the lack of systematic research on the effectiveness of policy measures, an increasing number of ad hoc decisions have been made with counterproductive effects on the availability of essential drugs. Future challenges demand new policies, for which recommendations are offered.

Conclusion A fertile ground for high-priced drugs has been created by changes in drug life-cycle dynamics, the unintended effects of patent legislation, government policy measures and orphan drug programs. There is an urgent need for regulatory reform to curtail prices and safeguard equitable access to innovative medicines.

Keywords

Drug pricing, Life Cycles, Drug Policy, Health care Costs, Patents, QALY, Equitable Access, Innovative Medicines, Essential Drugs

4.2.2 Introduction

Global health care expenditures have been rising sharply, and drug costs are a major factor. [1] Recent public outcry about exceptionally high prescription drug prices have made this subject a popular media and political topic. Discussion of drug prices has moved from an academic and government level to a broader society level and now includes the evaluation of public impact. The price of medicines was one of the campaign issues in the 2016 Presidential election in the US [2;3].

Many examples of high drug prices exist and are frequently discussed in the media. They include several types of therapeutic drugs and geographies. One often mentioned example is imatinib (brand name Gleevec®), a drug for chronic myeloid leukemia, which tripled in cost after the US Federal Drug Administration (FDA) allowed for a new indication. Novartis raised its price from $31,930 in 2005 to $118,000 per year in 2015 despite a huge increase in the volumes sold. The price hike occurred despite the fact that research costs for the new indication were included in the initial price. [4-6] Also in the US, the list price of sofosbuvir (Sovaldi®) is $84,000 for a 12-week treatment, or $1,000 a pill, [7] which has caused health plans to refuse routine coverage of this drug for hepatitis C virus (HCV) infection. [5;8] Sovaldi® alone accounted for 64% of US HCV-related spending in 2014, which totaled $12.3 billion. [9] Sovaldi® could be cost effective, since it prevents the ultimate need for a liver transplant, but the financial impact is too high for US insurance companies to make it available for all patients with HCV infections. [2;10;11] In Spain and Latvia, the cost of a complete treatment of Solvadi® was noted to be “unsustainable” by key stakeholders such as pharmacists and pharmaceutical policy experts. [12] The cost of an alternative combination of ledipasvir/sofosbuvir (Harvoni®), marketed by the same pharmaceutical company, Gilead Sciences, is comparable to

a course of Sovaldi®. [13] But the high prices for HCV drugs are not the exception. In another example, US patients suffering from cystic fibrosis were denied reimbursement for a new drug – ivacaftor (Kalydeco®), with an annual cost of $311,000[13;14] – unless their health worsened on older, cheaper treatments. [13] The cost of pyrimethamine (Daraprim®), a 60-year old drug, rose from $13.50 to $750 per pill (a 5455% raise) after Turing Pharmaceuticals acquired the distribution licence. This has further sparked public debate. [15-17]. Additional price hikes in Mylan's EpiPen® from $94 ten years ago to $609 for a pack of two have caused additional public backlash, protests and US Congressional hearings. [18-20] The results of a recent trial, [21] which show that 74 patients needed to be treated for two years with the new cholesterol-lowering Evolocumab (Repatha®) to prevent one cardiovascular event, indicate that with the current list price of $14,523 per year, the prevention of one event would cost over $2 million. [22] These and many more examples of high prices for medicines, however innovative, are untenable and frequently beyond the ability of individuals, health insurance companies or even governments in high-income countries to pay. [23]

Governments and health insurers are struggling with the dramatic increase in costs of new medications. [7;24-26] In December 2015, the US Senate issued a warning report on Sovaldi's escalating drug price and its impact on the US health care system. The committee report said the Gilead Sciences pharmaceutical company had set the price as a benchmark to "raise the price floor" for its future hepatitis C-drugs like Harvoni, thus knowingly reducing the number of eligible patients for these superior treatments to cure HCV. [9] US congressional committees have opened enquiries into similar drug-pricing practices. [27] Simultaneously, on the other side of the Atlantic, the UK cost gatekeeper, the National Institute for Health and Care Excellence (NICE), initially rejected reimbursement for two costly cancer immunotherapies — nivolumab (Opdivo®) and trastuzumab/emtansine (Kadcyla®) — despite fierce opposition by industry and patient groups. [28] In both cases, the costs were estimated to amount to £90,000 per patient per year. [29;30] With a number of better targeted immunotherapies – that fit within highly promising precision medicine approaches – on their way to the market, the drug pricing and funding crisis is expected to deepen and reach a critical level for even the wealthiest countries. [31] The German government is planning to curb companies' right to set launch-

prices. Belgium, Luxembourg and the Netherlands are working together to seek a common approach to their price negotiations with drug firms. A January 2017 Lancet commentary co-authored by the Dutch Minister of Health Edith Schippers stated that: "We need meaningful efforts by both the pharmaceutical industry and governments to invest in new medicines, provide full transparency on costs, prices, and who pays what beforehand, and respect the legal space for governments to protect public health. If we don't succeed in these efforts, we cannot guarantee people's access to innovative and affordable medicines". [32]

On average, countries in the Organization for Economic Co-operation and Development (OECD) spend 17% of their health care budgets on pharmaceuticals;[24;33] in some countries, this is even more. [25;34] For low- and middle-income countries (LMIC), drug expenditure can be a critical public health problem[35-38] with some drugs out of reach for even well-insured patients. [26;39] In some cases, to prevent striking increases in premiums or taxes, regulators are forced to limit access to healthcare, [13;24;40-42] which leaves patients without the best treatments. [43] Of concern, is that the pharmaceutical industry might be tempted to view these high-priced models as the direction for future drug pricing of new drugs that impact larger populations. [13;44-46]

The prescription drug price controversy is not new. In the 1990s, there were comparable heated debates on the high prices for interferons, paclitaxel (Taxol®) and HIV/AIDS medication. [1;47;48] Though the prices of these drugs were much lower than current new drug price levels, the fact that taxpayers had helped to pay for developing those innovative therapies at the time, generated public debate on fair pricing. In LMIC, where the need for HIV/AIDS medication was the highest, the fair-pricing issue was even more pressing, particularly with regard to the problematic availability of essential HIV medicines. [38;41]

Pharmaceutical expenditures are based on two factors: price and volume. This means that regulation can either aim to lower drug prices, or reduce usage. [34;49;50] On the one hand, there is a growing life expectancy (and aging population worldwide), while there are increasing medical options for disease control. [51;52] Therefore, following drug innovation expectations and usage growth statistics, it is likely that costs will continue to rise. [53;54] Many countries are striving towards universal health coverage, with guidance from the global public

community, [55-57] to reduce individual catastrophic spending. [58] Although these countries are preventing individual catastrophic spending by pooling risks and costs, a sustainable solution to the problem of fast-rising drug costs is still necessary. The solution will require unprecedented measures to prevent health care costs from spiraling out of control. [59]

Many articles have been written about the high cost of drugs. Most seek to define the cause of high drug prices in terms of government policies or industrial pricing strategies and propose related policy measures to combat the phenomenon. This review takes another angle, and presents a comprehensive analysis of the long-term dynamics of pharmaceutical markets, drug life cycles and the sometimes unintended, counterproductive effects of market interventions by governments and health insurers. The aim is to determine what has caused the recent exponential rise in drug prices, and what can be done in terms of measures around drug pricing to safeguard equitable access to innovative medicines.

The article is structured as follows. First, drug life-cycle dynamics are discussed. Next, government interventions aimed at reducing drug prices and their consequences are highlighted. Finally, we provide suggestions for alternative policy measures to reduce drug prices and improve access to innovative and essential medicines.

4.2.3 Method

The prescription drug price controversy has been developing for several years now, but 2015 brought a significant change in public perceptions. Several new expensive drugs were introduced, and even old drugs were subject to price hikes. To find causes and possible solutions for recent price increases, a systematic review was performed.

Due to the wide selection of journals, PubMed was used as the search engine for peer-reviewed scientific articles. The PubMed search was performed on February 24, 2017. The search strategy was (Drugs OR medicines) AND (prices OR costs) NOT (Cost-effectiveness OR Clinical Trial OR treatment), published between January 1, 2014 and January 1, 2017, in English, and with full text available in PubMed. The articles were screened by TvdG based on the title and abstract to determine inclusion, and then read in full. Additional scientific articles were included based on the reference lists of selected journals since due

to an inherent time-delay, systematically searching for topic-related articles in scientific peer-reviewed journals provided insufficient coverage of the controversy.

To include the most recent developments, additional searches in the databases of a select number of reputable English-edition international financial and daily online newspapers were performed. For this purpose, the Financial Times (FT), the New York Times (NYT), the Guardian and the Wall Street Journal (WSJ) were selected to ensure insight from both an American and European perspective, and from both financially focused and general newspapers.

The newspaper articles were limited to those published between January 1, 2015 and December 31, 2016. For the FT and the NYT the terms "Drug AND Pricing" were used in the LexisNexis search engine due to the availability of articles. The Guardian was searched for the exact combination "drug prices" using the Google-based search engine on their website. The WSJ was searched with ProQuest, with "Prescription AND Drug AND Prices" as search terms. The news articles selected by title only were all read and selectively added until there was a saturation of citations and information. The result of the searches and selection procedure is displayed in Fig 1. To ensure the quality of reporting, the Prisma checklist was used. [60] This study was not registered with PROSPERO.

Inclusion criteria

- Published in peer-reviewed journal or selected newspaper
- Published between January 1, 2014 and December 31, 2016 for scientific articles
- Published between January 1, 2015 and December 31, 2016 for newspaper articles
- Published in English

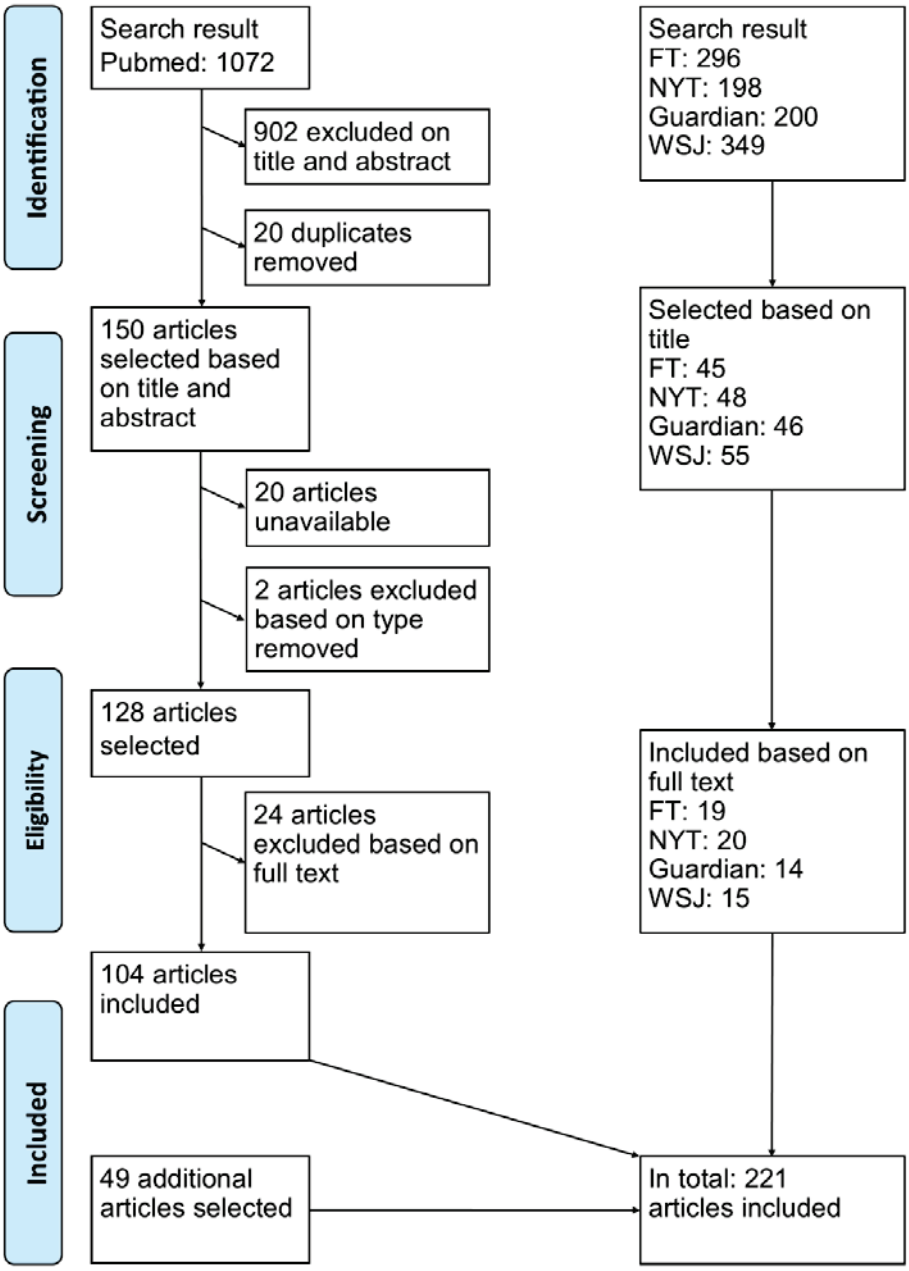

Fig 1. PRISMA flow chart. Schematic overview of the study selection process.
FT: Financial Times, NYT: New York Times, WSJ: Wall Street Journal

4.2.4 Life Cycles and Market Dynamics

The justification for high prices for pharmaceuticals can be seen as part of the changing nature of drug life cycles and market dynamics. Further details on both these aspects are presented in this section.

Life cycles describe the market behavior of many products. Generally, the product life-cycle pattern is represented by a "bell shaped" graph, a parabola, as exhibited in Fig 2. Though specifics can vary wildly, the general shape of the curve of investments during the drug development phase, exponential growth of sales after registration and decline through competition and patent term expiration is valid for most drugs. [61] Drug life cycles generally have four stages. First, there is a testing and approval trajectory. Second, after the drug is introduced there is market expansion, and the product is accompanied by growing expectations and drug indication extension. Next, drug maturity with a high sales volume is accompanied by rising criticism and disappointment regarding drug effectiveness and side-effects. Finally, there is contracting use and limited drug application. In most cases, this is a gradual process that involves the documentation of less favorable experiences and reports of the drug's effectiveness and adverse reactions in everyday practice. Thus, a drug's benefit-risk assessment and the resulting safety profile is under constant revision. Over time, newer and presumably better alternatives gain attention. This is part of an evolutionary process of selection and adaptation. Most brand-name medicines continue their careers as generics after their patents expire. On average this results in a 20-25 year therapeutic life-time in 'the doctor's bag' – the portfolio of drugs available to a doctor – due to therapeutic substitution and competition between branded drugs and generics. [62]

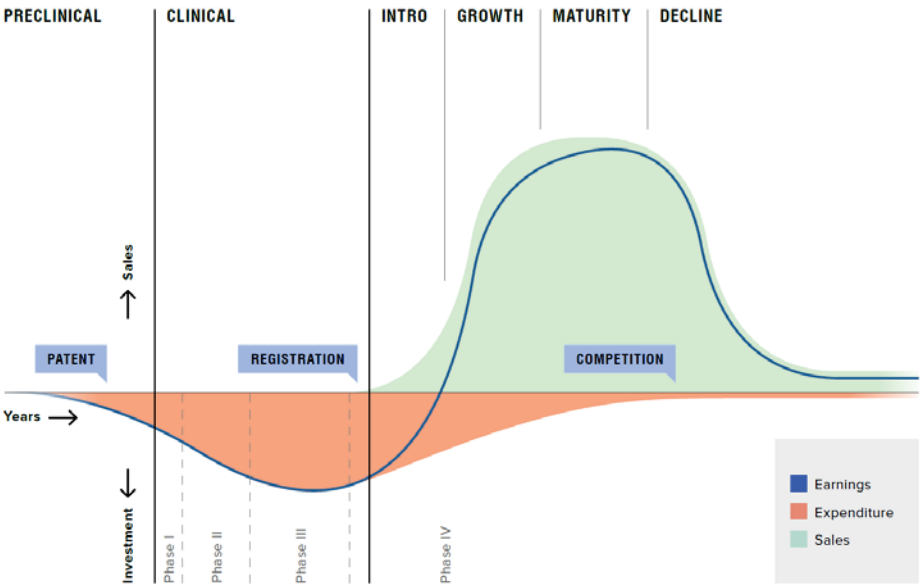

Fig 2. Drug life cycle curve. General curve describing an innovator drug's investments and earnings during R&D and market performance. The life cycle phases are indicated above the graph, and the phases of the R&D trajectory are below the graph. Own work.

What we tend to forget is that therapeutic drugs are inextricably linked to a dynamic mixture of opinions, practices and rituals and as such are important tokens of healing as a cultural process. Doctors, regulators, health care insurers, patients and politicians, however, prefer to believe that it is evidence-based medicine and not enthusiasm or belief about a particular drug that makes a drug more or less therapeutically effective. Dan Ariely, a behavioral economist at the Massachusetts Institute of Technology, and his research team have shown that a 10-cent pill doesn't kill pain as well as a $2.50 pill, even when they are identical placebos. [63] This may partly explain doctors' and patients' preferences for high-cost brand-named drugs over inexpensive, widely available, chemical and therapeutic generic equivalents. [64]

4.2.4.1 Types of prescription drugs

This article refers to prescription drug prices, but there are distinct types of prescription drugs and this requires clarification. First, there are drugs that are under patent, with an exclusive producer and no direct competition. Then, there are generic drugs with an expired patent that allows for production by other manufacturers.

Biological drugs follow the life-cycle patterns of small molecules or conventional drugs, but higher prices are accepted and specific regulation of generic competition is in place. Oncological drugs are a separate category, because high prices are historically more common, [26] expected to rise, [5] and more acceptable given the severity of the indications. Laws are in place to incentivize the development and marketing of orphan drugs, which means they follow market dynamics that differ from conventional drugs. Finally, when the patent runs out, and other producers can manufacture the same drug, generics are introduced. In the case of biologicals, biosimilars compete with the innovator while following a specific set of regulations. [65]

4.2.4.2 Patents and registration

The pharmaceutical industry is often characterized as a competitive sector in a free market, where the total supply and demand determine market price. However, according to business analysts, in a truly free and competitive market without patent regulation, it would be difficult to profit from new drug development. [66] This is why governments protect companies from competition during the life of a patent. In general, the term of a new patent is 20 years from the date on which the application for the patent was filed. [1;40;67;68] This can be extended to 25 years. [69] In addition, in the US, the FDA can grant exclusive marketing rights upon a drug's approval, which is generally concurrent with the length of a patent. The FDA usually grants new drug exclusivity for between seven years for orphan drugs[70] and five years for new chemicals, with an additional period of six months of exclusivity following pediatric approval. [71;72]

Patents are also granted for new chemical entities. This allows companies to charge high prices once the drug is ready for marketing. [1;73] Patents then become public, which gives other producers the chance to further improve and develop the drug. [41]

Patent timelines are limited, which provides an incentive for companies to shorten the drug development phase or look for disease areas with less stringent trial requirements. For example, there is more research in drugs for late-stage cancer than early-stage cancer, because of the less demanding and shorter trial trajectories. [1;74;75]

The number of patents a company files, or alternatively the research and development (R&D) costs per patent filed, are often used as

an output measure for the efficiency of drug development and the future of a firm. [76] Since most patented molecules do not make it to the market as an actual medicine, both datasets are incomplete representations of productivity. [77]

In debating the patent system, some analysts state that basic human rights like health and access to essential medicines should be equitable[23] and should not be limited by property rights. [1;6;78;79] Others use a utilitarian stance to argue that pharmaceutical companies are for-profit entities, [40;73] and without patents these companies would not be incentivized to develop drugs. [1] This difference in viewpoint is illustrated by the litigation surrounding patents and compulsory licensing (see paragraph 4.2.7) in LMIC. [73;80;81]

4.2.4.3 Developmental phase and registration

Pharmaceutical companies must register new drugs, which requires clinical studies and safety tests. This is a high-risk, high-cost and low-output endeavor. The odds of having a drug approved varies from approximately 24% (for systemic anti-infective drugs) to less than 10% (for drugs used to treat cardiovascular, gastrointestinal or metabolic disorders). [82] On average, it takes a company ten years to register a drug. [41;68;77] Thus, companies have to decide on projects that have a good chance of becoming registered drugs several years in the future. [83] The drug development process requires investments, estimated at between $60 million to $2.6 billion, [6;67;68;77] though most estimations are close to $800 million from bench research to prescription medicine. [33;66;72;84;85] The wide range of cost estimates is due to the lack of clear data and various methods of calculation, and depends on the type of drug and the trial data required, [72] as well as the size of the company developing the drug. [86] Development costs are highest for large companies due to their relatively high overhead and marketing costs. [1;76]

Historic examples illustrate what happens when the demonstration of medicine safety during development is not adequately regulated. An exemplary case is the thalidomide drug disaster that took place between 1958 and 1962. [24;71;74] This drug for morning sickness resulted in malformations in the extremities (phocomelia syndrome) of thousands of babies born to women who had taken thalidomide during pregnancy. [71] Regulatory reaction to drug safety alerts often involves the introduction of more stringent regulations requiring more safety and

efficacy studies, which leads to more dropouts in the development process and an increase in invested time and costs. [41;74] Regulatory agencies are criticized by many parties for being either too stringent (delaying innovation and increasing costs) or not stringent enough (allowing dangerous drugs to be marketed). [72] Arthur Daemmrich, a US historian, discussed this tension between safety management and drug innovation and was the first to use the term 'double bind trade-off phenomenon'. [87]

The imperative of regulation makes it more difficult for smaller companies to register drugs, thus limiting the number of firms with the critical mass and financial means to invest in drug research. This situation limits viable competition from smaller companies [41] for Big Pharma--the collective sector of large pharmaceutical companies. That is why most new drugs that received a positive reaction from the Committee for Medicinal Products for Human Use (CHMP) of the European Medicines Agency (EMA) between 2010 and 2012 were filed by large (59%) or intermediate-sized (28%) companies. [88] Small enterprises are important during early phases of development. However, in later phases, if the success of a new chemical entity developed by a small company is likely, a large pharma company will often buy the small company or purchase the licence for the new medicine. [88]

R&D for medicines has been declining. [41;67;68;89] Higher investments, however, will not necessarily fill R&D pipelines with new promising drug compounds. R&D has recently yielded fewer drugs than in years past, since low-hanging fruits have already been harvested. [66;89;90] Furthermore, there are many drugs with promising results in phase II settings that have not made it to phase III settings.

Regulatory agencies allow drugs to be released to the market based on safety and effectivity, but not with reference to price or cost-effectiveness. [70;91] This means the price and reimbursement of a drug are determined only after registration approval and insurance company and/or government negotiation. [91]

4.2.4.4 Post registration and reimbursement

Once a drug is registered for a specific disease indication, manufacturers can apply for reimbursement. Many public health care systems allow the government to control drug prices. Some base the acceptability of a price on the Incremental Cost Effectiveness Ratio (ICER) and budget impact.

[10;83;92] This means companies have to assess the volume of sales [93] and the price at which they are reimbursed, and then offer a price based on that estimate. [90;94] Then, negotiations take place between the company and the reimbursing agent or government to determine an acceptable price for each stakeholder.

A drug's reimbursed price can be lower than the pharmacy retail price or list price. This makes patients aware of drug prices, since they will have to pay for the difference out of pocket. Such pricing and reimbursement schemes can be a tool to make patients switch to cheaper or generic drugs, and make manufacturers of high-priced drugs lower their prices to prevent patients from making this switch. [36;59;95] Manufacturers argue that patient co-payments can cause adherence problems, [96;97] especially for expensive [6;98] and psychiatric drugs. [99] This means physicians and patients prefer drugs without co-payments. [78;100] To circumvent this situation, producers have implemented patient-assistance programs, which are discussed in paragraph 5.6.

Companies want to make the highest possible profits in each country by differentiating prices, [40] but they also want their prices to be similar across countries and close to competitors to reduce the incentive for parallel importation. [24;101;102] Governments worldwide want innovative new drugs to be available as quickly as possible, so their population can profit from them. High drug prices may incentivize companies to develop and launch their new drugs faster. [1;40;73] On the other hand governments also want to have affordable drugs for everyone at the lowest possible price, to reduce healthcare spending. [1;5;78]

In the US, the government does not control reimbursement, [16;33;91] based on the assumption that the free market will drive the pharmaceutical industry to compete, which will result in lower prices. [92;103-105] Thus, pharmaceutical companies set their own prices, which allows for market calculations aimed at maximizing profits. [106-108] As a consequence, US prescription drug prices are among the highest in the world. [6;33;109] For the uninsured, a cancer diagnosis is still a major cause of personal bankruptcy. [110;111]

In general, the manner in which drug list prices are determined is not transparent, [14;57;106;112] so critics are pushing for more

transparency. [77;98;113] Some differences in pricing between countries can be explained by differences in health care systems [4;93;114] and socio-economic dynamics, [78;84] all of which lead to differences in ICERs. In some cases, a drug's price is based on the old standard of care plus a premium, [5;6] the uncertainty associated with the drug [93] or simply the government's willingness to pay. [92;115] Ultimately, a drug's price is unrelated to the cost of development [4;5;51] or a country's gross domestic product, [12] and only related to cost-effectiveness if the payer introduces this into negotiations. [30;53;107]

4.2.4.5 Introduction and growth phase

When a drug is marketed, there is a level of market penetration. [61] It is important to the producer that the drug is brought into use as quickly as possible, since a patent limits the period of exclusivity and thus profit. Speeding up market penetration was historically accomplished by advertising to doctors, pharmacists and patients (see Fig 3). [116] In the US, doctors are paid directly for promotional activities, [112] though the Affordable Care Act (ACA) requires disclosure.

Direct-to-consumer advertising (DTCA) is controversial, since it increases demand, partially through inappropriate prescription. Doctors will prescribe drugs that are not indicated if they know they can satisfy their patients by doing so, or unnecessarily prescribe branded drugs. This increases costs without improving health. [117] Therefore, this practice has become illegal in most countries, but not in the US or New Zealand. [1;117-119] This situation makes it harder for drug companies to create demand for their products in most countries, and reduces turnover in the early phase of developing a new drug.

In Europe, marketing to doctors and pharmacists is permitted, if it is medically substantiated. [120] This requires more expensive studies, and careful wording of the marketing message. Still, marketing is a large part of the pharmaceutical industry's expenses. In fact, more money is spent on marketing than on R&D. [24;41;121]

To market drugs to doctors and circumvent this regulation, trials are sometimes used as a marketing tool. [48] 'Seeding trials' are designed to 'seed' the use of a drug among patients and physicians, while they often offer no scientific purpose. [122;123] Furthermore, medical centres that want to promote an image of being at the cutting edge of science are willing to test the newest available experimental treatments

and compete for hosting profitable industry-sponsored clinical trials. [48] This was the case for interferons when companies were looking for an indication for this new experimental drug. [48] Numerous promising new drug compounds have followed suit. [122;123]

A pharmaceutical company's marketing push is somewhat mitigated by the increased tendency to work with local guidelines. [25;50;73;124] This often improves the quality of care for the patients, but makes it harder for new drugs to enter the doctor's bag. In addition to the cycle of updating those guidelines – which generally happens once every five years, or less frequently [46] –new drugs have to show superiority to already available drugs. Furthermore, the price of a new drug is often higher than that of older drugs that might be available as generics, which increases the demand for proven superiority. [41]

4.2.4.6 Mature phase

During a drug's patent life, doctors and pharmacists play a crucial role in the choice for one drug over another. [1] These professionals need to inform the patient about their pharmaceutical options, and a drug's effectiveness and costs. This is why advertising aimed at brand recognition continues during the mature phase of a drug's life cycle. [61]

Several studies have shown that if there are financial incentives for doctors to choose one drug over another, the one that is most beneficial to the doctor's finances is most likely to be prescribed. [1;125] In the US, where Medicare and Medicaid reimbursement are based on a 6% mark-up of the price of cancer drugs, [53;110;119] doctors have an incentive to select the more expensive option. [1;5] This is another explanation for the high prices for drugs in the US. In order to help patient and doctors, the European Society for Medical Oncology and the American Society of Clinical Oncology have developed frameworks to assess the value of new cancer drugs. [58;126]

To combat incentives to prescribe expensive drugs, some health insurance companies in the US offer monthly payments of $350 to physicians who prescribe according to their guidelines, thus saving costs and improving patient health. [5] This scheme compensates a physician for earnings missed by prescribing less expensive drugs. [5]

In this phase, companies often attempt to have their drug registered for additional indications, thus increasing the number of patients, to increase their sales volume. A larger patient base would

logically make the cost per treatment lower, but this is often not the case. [53;111]

4.2.4.7 Declining phase

Several countries around the world have implemented preference policies, aimed at generic substitution. [1;127;128] This policy requires physicians to prescribe and pharmacists to dispense the cheapest available version of a drug, often generic, unless a more expensive one is medically necessary. This can be the case for drugs with a very small therapeutic window, like Tegretol® (containing generically available carbamazepine) for the treatment of epileptic seizures. [127] For these drugs, the preference policy implies that new patients start on a generic drug, but those who have already reacted well to a branded version do not have to switch. The potential substitution rate differs per indication group. [103;127;129;130]

Generic substitution reduces the length of the life cycle of patented innovator drugs, since they are replaced by generics as soon as the patent expires. The time an innovator drug has to generate income is thus reduced, which makes higher prices during the exclusive phase necessary to generate the same turnover.

Doctors and patients are often hesitant to change the brand of the drug after patent expiry, since they are familiar with the innovator and trust it. [69;96;99;128;131-133] This explains why generic substitution only happens on a large scale if policy to enforce it is in place. In some countries, generic substitution by pharmacist's initiative is mandatory; [78;130] in others it is prohibited. [99;130]

Since roughly 2000, Big Pharma has been struggling with the patent cliff, a series of blockbuster drugs whose patents have expired. This has caused a significant loss of turnover due to generic substitution. [66;67] The effect cannot be compensated for by new drug introductions, since relatively few new blockbuster drugs have been introduced. [103] This means that in order to maintain profitability, more revenue must be generated from fewer breakthrough drugs, which has led to increased prices for innovator drugs and increased merger and acquisition activity within the pharma industry. Over the last two decades, 60 pharma companies have merged into just ten pharma companies. This consolidation has helped Big Pharma gain more power to influence

regulation and pricing policies while simultaneously diminishing competition (see Fig 3). [134]

The patent expiry and substitution effect is expected to be smaller for the new generation of biologicals, because these biological drugs are more difficult to copy. Also, a drop in price in cases of biosimilar substitution is often just 20-30%, [65;85;135] whereas the price for conventional small molecules can fall by more than 70%. [85;99;131]

Large pharmaceutical companies employ several tactics to extend the life cycles of their products, and reduce their loss of income due to patent expiry. Strategies to address this often include improved formulations (e.g. Seroquel® XR, quetiapine), new indications (e.g. Zyban®, bupropion), chiral switching (e.g. omeprazole (Prilosec®)/esomeprazole (Nexium®) or citalopram (Celexa®)/escitalopram (Lexapro®)), combining drugs (e.g. Harvoni®, containing ledipasvir and sofosbuvir), changing to over-the-counter (e.g. Prilosec®, omeprazole) and finally introducing an authorized generic[6] (e.g. atorvastatin, Lipitor®/Zarator®). [67]

Another option is the development of pediatric dosage recommendations and formulations, which allows for extension of market exclusivity by the FDA for six months. [71;72] Finally, 'pay for delay', the act of paying off generic competition so an innovator can maintain market exclusivity, is fiercely contested by both the European Commission and US Federal Trade Commission, but the practice has been increasing over the last few years. [67]

Some drug brands are so strong that, even after the loss of market exclusivity, doctors and patients continue to privilege them over generic drugs. Examples include brands like Viagra®, Prozac® and Aspirin®. [136] For over-the-counter medicines, in particular, branding is a relevant mechanism to maintain market share, since consumer name recognition is a more important factor in product choice when there is no medical professional role.

Fierce competition between generics can cause companies to offer prices that become too low to be sustained by the offering company. Sustainability can be threatened by the rising prices of raw materials and production costs, [137;138] which lead to a failure of supply and shortages. [95;131;138] This situation can also force competitors for a

specific drug to completely retreat from the market, which leads to fewer producers[139] or even a new monopolistic position for a generic, [137;140] as recently seen with Lanoxin® (digoxin) and Daraprim® (pyrimethamine). The result is that even for generic medicines we increasingly see steep price increases. [141]

4.2.4.8 Changing life cycle dynamics, effects on drug pricing and profitability

Drug life cycle analysis indicates a trend of shortening life cycles and pharmaceutical companies experiencing more difficulty achieving high, sustainable sale volumes during the past two decades than before. Since a company's income is based on volumes multiplied by price (equals value), the first strategy to maintain high revenues is to increase price. [142] Despite regulated pricing, this practice results in drug spending growth matching overall medical spending growth. [39]

On average, the top ten pharmaceutical companies have a profit margin of 20%;[2] those noted in the S&P 1500 have a net profit margin of 16%, compared to 7% for all other companies in the index. [119] This means that even though companies experience more difficulty in achieving long-term high-volume prescription drug sales, the higher drug prices compensate for the lower product turnover and safeguard Big Pharma's high-profit profile. This is not surprising, because pharmaceutical companies are for-profit entities that wish to maximize their profits and increase share-holder value without breaking the law. [24;40;143] However, this approach means they may not automatically do what is best for society. Critics argue that more regulation is needed to counterbalance Big Pharma's only-for-profit motive and force them to do what is best for all stakeholders. [15;53;104;144;145] Through a number of interventions (some more effective than others) governments and their regulators have tried to direct either the price of drugs or the availability of innovations. Government interventions to stimulate or curtail the pharmaceutical markets and the introduction of new procedural measures concerning drug patent licences and drug registration licences are discussed in the next chapter.

4.2.5 Drug innovation, regulation and pricing interventions

As stated previously, though the pharmaceutical market is often portrayed as a competitive market, it is not truly a free market. In addition to the patent system, skewed economic dynamics create further complexities. In free markets, a consumer decides on, buys, pays for and

uses a product, whereas in healthcare, a doctor decides and the pharmacy or hospital pharmacy provides, the insurance company or government pays and the patient uses the product. [142;146] An overview of the stakeholders and their relationships is given in Fig 3. Financial incentives are not aligned with consumption, so companies' pricing power is not related to how consumers value the products.

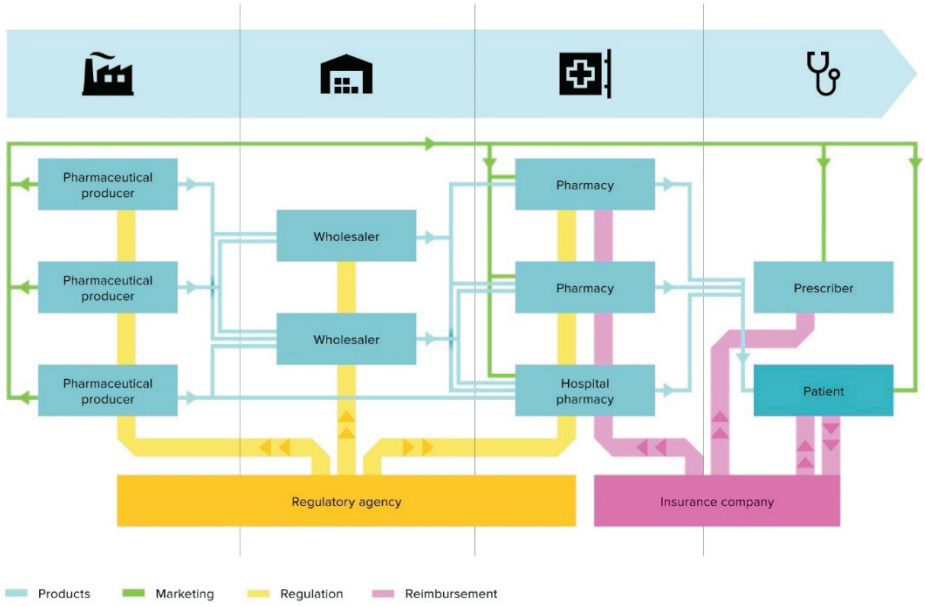

Fig 3. Stakeholders. A simplified, schematic overview of stakeholders and relationships in the pharmaceutical market. Own work.

4.2.5.1 Unintended consequences of innovation and measures to stimulate drug safety

Drug regulation usually aims at making new drugs available while keeping costs down, however, there can be unintended consequences. For example, if drug production or distribution becomes too competitive to remain lucrative, the market can become so consolidated that drugs are no longer available. [140] Even worse, temporary or sustained monopolistic positions can arise due to market failure which can cause an increase in prices. [104;106;147] Hence, after specifically designing a policy to stimulate innovation and safety or reducing prices, potential consequences should be carefully monitored.

4.2.5.1.1 Orphan and priority drug regulations and potential consequences

There are cost-reduction strategies that could work by changing the trajectory of a drug's development. [53;68] The first option is to speed up innovation and regulatory approval, so that companies have less waiting time before marketing a drug and thus enjoy a longer profit-generating post-marketing patent life. [5] One way to do this is to accept surrogate parameters as trial endpoints to prove efficacy, which saves time. [135] Another innovative option is to harmonize regulation between countries, so companies only have to prove efficacy once. [135] However, because drugs are not necessarily priced based on investment costs, the effects of this approach on drug pricing might not be significant, or as we will see, harmonized regulations may actually be counterproductive.

One clear example where regulation is in place to speed up innovation is the field of orphan drugs. Generally, low patient volumes make it unattractive for pharmaceutical companies to invest in the development of orphan drugs. [148-150] Fabry disease, for example, has a prevalence of approximately 1 per 100,000 persons, thus making it unattractive for companies to develop drugs for these patients without further incentives. [149] That is why the orphan drug regulation was designed.

The benefits of developing drugs with an orphan status are exclusive licensing for seven years, faster assessments and lower taxes in the US. European regulatory bodies offer exclusive licensing for ten years, lower regulatory fees and scientific advice. [148;150] Orphan drug legislation has worked, yielding at least 73 drugs for orphan indications in the European Union since the law passed in 2000[151] and 335 in the US since the FDA set regulations in 1983 and 2002. [152]

Another incentive for companies in the US that develop orphan drugs is a priority review voucher, which is also available for tropical disease drugs. This voucher is awarded after a company develops a drug for an orphan disease and releases the patent. When applied, it allows companies to request an expedited review process for a new drug, which can speed up the regulatory process by several months. This approach allows a pharmaceutical company to stretch the patent period, and thus the mature, beneficial part of the life cycle of a subsequent blockbuster drug. [153] A well-used voucher could increase a company's income by up to $300 million according to some estimates. [1;154] The company

can also sell the voucher to another company that is about to launch a similar drug for roughly the same price. [154]

The moral argument for orphan drug regulation is that society, and medical science in particular, has an obligation to pursue new therapies for everyone, including people who suffer from orphan diseases. [149;152] The downside of this policy is the non-utilitarian outcome that money is being used for diseases that very few people have, and that the same money could have been used for more relevant research reaching a larger population. [1;145;151;152]

Given that no price ceiling or maximum budget impact is imposed on an orphan drug's regulatory design, orphan drugs are often very expensive. Prices are unrelated to effectiveness or prevalence, which means that regulatory bodies frequently label these drugs as not cost-effective. [148;151] Also, the evidence for the effectiveness of orphan drugs is often lower in quality than required for regular drugs, and more side effects are tolerated than for other drugs. [148]

In Europe, market exclusivity can be withdrawn after five years if a product has generated adequate profit. However, this has not happened for a single orphan drug to date. [149;155] Even for orphan drugs that have lost exclusivity, no generic producer has ever created a competing product, and it is doubtful whether or not this would ever be attractive given the low patient numbers. [151;156]

Orphan drug regulation is an example of a policy working better than expected, thereby increasing healthcare budgets. Companies have tried to abuse this regulation for profitability purposes. For example, there was a request that malaria be designated for an orphan drug indication in Europe. [71]

Orphan drug regulation is also used by pharmaceutical companies to register drugs so that multiple indications – not always orphan indications – can be added. For example, Gleevec® (imatinib) has been marketed for several orphan indications, [149] yielding a turnover of $4.7 billion in 2012. [6] Rituximab, the world's second most profitable drug, also holds multiple orphan drug indications, in addition to the use for common rheumatoid arthritis. [155] Although this approach increases drug volumes, companies usually do not reduce prices. [152;157] Finally, pharmaceutical companies split common indications

into groups that are small enough for their drugs to qualify for an orphan drug label. [70;149;155]

Regulations for orphan drugs are quite effective, but should include compulsory price-ceiling measures to prevent tax-payers from paying twice, first for some of the R&D costs and second for reimbursement of overpriced drugs. Furthermore, the orphan drug label should only be applicable if a drug has not been granted access for another indication already. Currently, Revatio® (sildenafil, also known as Viagra®, for erectile dysfunction) is being sold as an orphan drug for pulmonary arterial hypertension. [149]

4.2.5.1.2 The FDA's unapproved drugs initiative and consequences

In June 2006, the FDA announced a new drug safety initiative with the goal of removing unapproved 'old' generic drugs with problematic safety profiles from the market. The FDA states that it uses a risk-based enforcement program in order to focus on products that pose the highest threat to public health and "without imposing undue burdens on consumers, or unnecessarily disrupting the market". [158] However, the program has had unintended consequences. If a product is not officially approved by the FDA, the agency can require a New Drug Application from the manufacturer, which is reviewed to determine if the drug meets FDA standards. Inexpensive generic drugs that have been on the market for decades are studied anew, drug applications are filed and exclusive patent rights to sell the drug are given to the first manufacturer who meets the new FDA effectiveness standard. This manufacturer can then decide what to charge — with no competition. Exemplary is the patent flipping of the formerly inexpensive drug colchicine (less than $1 for 30 pills), used to treat flare-ups of gout. Now, after FDA review, just one manufacturer has the patent rights to market colchicine as Colcrys®, and the retail price is almost $200 for 30 pills. How drug pricing measures could counteract these pricing strategies by the pharmaceutical industry is discussed below.

4.2.5.2 Possible drug-pricing measures

There are many ways to reduce spending on drugs. However, all are based on one of four general intervention options:[36]

a) Shift from expensive to cheap drugs, within the same class,
b) Shift costs towards patients or insurers,
c) Reduce drug prices,
d) Reduce total drug uses.

It is not clear which mechanism is the most effective, but authorities in many countries often implement policies encompassing several of the abovementioned options. An overview of possible regulations and the mechanisms through which they reduce costs are given in table 1.

Schemes to reduce drug prices are used most often to reduce overall healthcare spending. For example, according to one calculation setting prescription drug prices 20% lower than the current list price would increase the number of users who can afford the drug by 23%, while decreasing revenues for the drug company by only 1%. [33] More examples with comparable outcomes exist. [4] Of course, the outcome is completely dependent on specific market conditions, prices and regulations. The general consensus is that reducing prices increases the number of users, and this could at least partially offset losses due to lower pricing. A popular argument against paying less for drugs is that innovation would not be financially worthwhile, and society would not enjoy the possible benefits of new innovator drugs. [77] This argument will be discussed later on.

Table 1. Policy effects. This table lists the policies that are in effect in various parts of the world, their effects and their unintended consequences. EU: European Union, USA: United States of America, DTCA: direct-to-consumer advertising, UK: United Kingdom, LMIC: Low- and Middle-Income Countries.

Policy	Location	Mechanisms	Effects	Side effects
Patent laws	Worldwide	Increase innovation	Gives an incentive for innovation	Increased prices during patented period, reduced transparency in research
Orphan drugs	EU, USA	Increase innovation	Many new drugs have been developed	High prices, low quality evidence of effect
Biosimilars	EU, USA	Shift drugs	Availability of generic versions of biological drugs	Concerns about comparable effectivity and rare side effects due to fast market authorisation
Development cost reduction	EU, USA	Increase innovation, reduce price?	Faster entry of new drugs	Lower quality evidence

Limiting DTCA	EU	Reduce use, shift drugs	Reduced inappropriate prescriptions	Reduced awareness of new drugs for professionals and the public
Reference pricing	Several EU countries, Canada, Australia	Reduce price	Lower prices	Best payers get drugs first
Price ceilings	Several EU countries, Canada	Reduce price	Fewer price increases	Higher initial prices
Value-based pricing	Several EU countries	Reduce price, shift drugs	Decision making is more evidence-based, and treatments are rewarded for actual efficacy	Prices are set just below cost-effectiveness threshold, limitation of options
Preference policy, compulsory generic prescribing	Several EU countries	Shift drugs, shift payer, reduce price	Increased use of generic drugs	Shorter life cycle of patented drugs
Stimulate guideline adherence, pay for performance	EU, USA	Shift drugs, reduce price	Better prices and quality of healthcare	Limited options for treatment
Negotiation power through monopsony	New Zealand	Reduce price, shift payer	Reduced prices for population through increased negotiating power	Shift of costs to countries with less bargaining power
Voluntary out-licensing	LMIC	Reduce price, shift drugs	Lower prices in LMIC	Counterfeit parallel trade
Open tenders for exclusivity	Several EU countries, Russia	Reduce price, shift drugs	Reduces prices	Drug shortages due to less dynamic supply chains
Compulsory licensing	LMIC	Reduce price, shift drugs	Makes drugs affordable for LMIC	Counterfeit parallel trade
Incentivize physicians and pharmacists	Several EU countries, USA	Shift drugs, reduce price, reduce use	Directs towards prescribing cheaper drugs	Patients might lose access to more expensive brands
Profit limitation	UK	Reduce price	Lower profit margins, through lower prices or higher investments	Could incentivise companies to spend on less relevant causes

4.2.5.2.1 Biosimilar substitution regulation and resistance to substitution

Both the EMA (since 2003) and the FDA (since 2010) have regulations for accepting biosimilars to bring down the price of treatments by

increasing competition without reducing safety. [65;72;159;160] However, unlike the EMA, the FDA only allows for an interchangeability label if the manufacturer has shown that a biosimilar drug has the same effect and safety as the originator or for switching between them. [65] The regulatory framework for biosimilars is designed with similar intent as regulation for small molecule generic drugs, but its effect is thought to be less significant. This is due to a smaller price difference between the originator drug and generics, [143;161] as previously mentioned. Research and production costs for designing a biosimilar are significantly higher than for designing small molecule generics. [65;135;161] Thus, innovators can prevent biosimilars market penetration by offering discounts on the original biological. [135]

Furthermore, ongoing controversy on the interchangeability between biosimilars and originator drugs makes doctors and patients wary of using biosimilars. [159;162] It requires significant effort on the part of reimbursement authorities to overcome this unexpected and rather persistent unease about and resistance to the use of biosimilars, that is exploited by marketers of the biological originator companies. [159]

Pharmaceutical policy is often designed based on negotiations among various stakeholders, but doctors are not always invited to the table, according to critics. [163] If the medical community were more involved in creating biosimilar regulations and substitution programs, the effectiveness of regulatory and cost-reduction policies might improve significantly. [159]

4.2.5.2.2 Engaging physicians and pharmacists in price reduction programs

As stated previously, physicians and pharmacists have a central role in determining which patient receives which medicine, and whether the use of expensive drugs is beneficial for specific patients. [133;163;164] Programs that provide financial incentives for prescribers to save on costs incentivize physicians to be cognizant of drug prices and have the potential to reduce pharmaceutical expenditure gradually and permanently, by either rewarding when expenses are low or enforcing penalties when expenses exceed indicative or earlier budgets. [34;49;165] After the implementation of such programs, doctors are more inclined to believe that medical costs are a relevant consideration in drug usage. [166]

Just educating medical staff on drug pricing does not have a lasting effect. To change physicians' attitudes, the prices of drugs must be considered, and constant reinforcement and easily available information is necessary. In one example, simply adding a sticker to indicate the prices of anesthetic drugs per hour in operating rooms significantly reduced their use and therefore, costs. [167]

Another incentive to reduce spending is index pricing, which is similar to internal reference pricing. Drugs are classified in index groups of therapeutically interchangeable drugs. The prices for each group are determined based on the average drug used over the last period, which is frequently updated. The pharmacist is reimbursed for the price that is given for that index group, regardless of the drug that is actually dispensed. This incentivizes pharmacists to dispense the cheapest version of a drug, preferably below the index price so the difference can be kept as profit. This approach drives down the index price for the next period, thus creating a downward spiral. [36] Downsides of internal reference pricing are discussed in the next paragraph.

Finally, another way to reduce drug prices is for doctors to prescribe using the international non-proprietary names (INN) of drugs, and the brand names only when a brand is strictly necessary. [130] It is left to the pharmacist to dispense the drug based on the INN, so doctors have no financial incentives to prescribe more or specific brands of drugs. [100;164]

4.2.5.2.3 Reference pricing approaches

Reference pricing is a tool to set a benchmark for reimbursements. [36;95;168] Of all price control measures, this is the one with the most evidence for effectively reducing drug prices. [36;99] Reference pricing can be done in two ways: internally or externally.

Internal, therapeutic or national reference pricing is based on comparing a drug to other drugs with the same active ingredient or with comparable clinical effects within a country. The maximum price of the new drug is then based on the average or lowest price in that cluster. [1;55;99;163;168] This incentivizes companies to develop drugs for indications with no competition, particularly no generic competition, because that would bring the price down. [1;92] However, internal referencing discourages development in existing drug classes. [1] Late me-too drugs could be placed in a cheap class with therapeutic

equivalents. [55;169] This method of reference pricing only works by reducing prices and shifting patients towards cheaper drugs. [36]

External, or international reference pricing is used by most EU member countries. [78] It is based on comparing a new drug's price with other countries with a comparable economic status, [36;59;170] or with differing economic status after which the outcome is adjusted to purchasing power parity. [171] The mechanism of pricing used by a specific country, such as value-based pricing, can also be a reason to refer to that country. [55]

For example, Norway reviews nine countries and takes the average price of the three lowest prices. [119] These prices are then regularly updated. [99] This approach often provides a skewed picture, because it does not take into account undisclosed prescription drug rebates and discounts that most purchasers receive for prices on the official list. [78;99;126;172] This mechanism incentivizes companies to register in countries with the highest willingness to pay first. [155;170;172] Countries with a lower willingness to pay could see a drug launch delayed, or no launch at all, to prevent other countries from adjusting their prices downwards. [92;99;170;173]

4.2.5.2.4 Value-based pricing measures

The ideal pricing model should include the health and socio-economic benefits of a drug by deploying sophisticated out-come based compensation models. [105;174] The price of a drug should be proportionate to the added value in terms of quality of life, life years saved or tumor shrinkage. [6;24;95] This would improve the value per monetary unit spent on health care, and increase innovation in relevant areas. [90;175]

Currently, cost-effectiveness analysis is used in many countries as a factor in the pricing of new drugs, [78;176] but often drugs and treatments that are not cost-effective are still reimbursed, and sometimes by specific reimbursement funds like the Cancer Drug Fund in England. [177] Vice versa, cost-effective drugs are sometimes not reimbursed at all. [178;179] A major reason for this is the lack of standardization in the practice of value-based pricing. Which factors are included and which are not varies, so value-based pricing is currently more of an art than a science. [24] Data about the effect of such schemes are contradictory. [99] One factor is that this policy has given a perverse

incentive to drug companies to set high drug prices for the new generation of innovator medicines that are in line with the cost-effectiveness threshold (mostly in terms of quality-adjusted life year [QALY] and/or incremental cost-effectiveness ratio [ICER] terms) that a country is willing to pay. This also explains the differences in prices in individual countries, because cost-effectiveness thresholds differ across countries.

While not an official cut-off, the threshold for cost-effectiveness is set by the British regulators at approximately £30,000 per QALY, [155;181;182] and £50,000 for end-of-life drugs. [177;179] Other European countries use cost-effectiveness thresholds that can vary between €10,000 and €50,000 per QALY. [6] After extensive analysis, the total British health care system was found to provide care at £13,000 per QALY, which was much cheaper than many drugs. [181;182] This means that money would be spent more effectively on parts of the system other than high-priced innovator drugs. [177] Thus, the costs of high drug expenditures place a heavy burden on the health care system and may displace other high-quality healthcare services. [178;181;183]

Another practical difficulty in value-based pricing is a drug's indication. For example, Tarceva® (erlotinib) is more effective for lung cancer (extends survival by 3.5 months longer than chemotherapy) than for pancreatic cancer (extended survival by two weeks versus placebo). Although the price is the same for both indications, Tarceva® clearly creates less value for pancreatic cancer patients. [110;180] Italy has an indication-specific pricing system to address this. Other EU member states and the US, struggle to implement a similar system. [115;180]

Some ways to ensure that a drug is priced according to its 'true' value are: risk-sharing pricing, pay-for-performance pricing or outcome-based pricing. These pricing schemes allow governments and companies to renegotiate a drug's price based on its real-life performance in terms of effectiveness, sale volumes or cost-effectiveness. [56;99;185] In this way, companies have an interest in the real-life therapy outcomes (in terms of effectiveness), and not just the drug's performance in clinical trials (in terms of efficacy). For example, companies could offer a drug at a reduced (or no) cost and receive annuity-style payments if a drug reduces hospitalizations after approval. [2;169;180;186] Examples of outcome-based pricing exist, though only for specific indications and small populations. [169;180] Tracking the results in the real world is also

a difficult administrative assignment. [187] However, this scheme has recently been implemented for several larger indications, such as for the effectiveness of a new cholesterol-lowering drug [187;188] and for the hepatitis drug Sovaldi® in Japan based on volume. [189]

4.2.5.2.5 Setting price and profit ceilings

Another method of controlling drug pricing is to set price ceilings in various forms. For example, to combat the high prices of generic drugs in Canada, the government has recently negotiated a fixed price ceiling for six of the most used generic drugs. [190] This one-size-fits-all approach might still result in overpricing for some of the six, and be too low to supply the entire market for others. A lower price could probably be negotiated through alternative tactics, like an open-tender invitation, but the several Canadian states failed to agree on an alliance for bulk purchasing. [190]

In the UK, the government signed an agreement with the pharmaceutical industry that limits increases in spending on branded medicines to below 2% per year. If more is spent, the industry has to reimburse the government. [114;119;183] Companies that did not sign this agreement are subject to direct price control. [114]

One more option is to introduce upward price rigidity, by prohibiting increases in drug prices. [93;119] In the US, it is industry practice to increase the list prices of marketed drugs at least yearly by substantial amounts, synchronized with the competition. [146;191;192] Canada, however, only allows drug prices to rise with inflation. [119] This causes companies to set high initial prices, especially when the sales volume is uncertain. [93]

A rather alternate approach is limiting companies' ability to make high profits. An example of this is the Pharmaceutical Price Regulation Scheme (PPRS) in the UK. [36] If profits exceed a percentage agreed on after negotiations, a company must reduce prices, delay price increases or repay the excess to the Department of Health. [36]

This approach is designed to be an incentive for pharmaceutical companies to either reduce the prices of their drugs, or invest more of their income in research. Conversely, it could also incentivize companies to spend more money on areas that are not relevant to reduce profit and avoid paying the Department of Health.

Partnerships between companies and charity foundations can help direct pharmaceutical research towards a needy cause. For example, the GlaxoSmithKline–PATH Malaria Vaccine Initiative partnership yielded Mosquirix, the first malaria vaccine. [193;194] The partnership included a payment of $200m from the PATH Malaria Vaccine Initiative, backed by the Bill and Melinda Gates Foundation, to help fund pediatric trials. [195] The price of the drug was agreed to be the costs plus 5%, which is the same price as an insecticide-treated bed net, so that it would be available for those in need. [193;194] The profit is then reinvested in more malaria research. [194]

4.2.5.2.6 Lowering prices through open-tender invitation and negotiations

Many countries use open tenders to lower prices. [55] They invite manufacturers or wholesalers[55] to propose a confidential price for which they can guarantee to exclusively deliver a specific drug (e.g. simvastatin) or a drug in the same class (e.g. any statin) for the entire market for a term (often two years). The supplier or suppliers who offer the lowest price receive the contract, and secure income during the contract period. [99]

This approach effectively reduces drug prices, since generic drug manufacturers are stimulated to offer the lowest, but still profitable price. However, open tenders require a market size that is significant enough for companies to be interested, enough generic producers to create competition[50] and an agreement on the contract conditions. [190]

A downside to this market exclusivity for generics is the risk of shortages if the winning company fails to supply. [161;190] For companies losing a bid, it could be difficult to offer a competitive price in the next round due to loss of capacity, which reduces competition and creates new monopolies. [55]

Drug prices are often negotiated between national governments and companies, based on a reference price or a figure based on a drug's cost-effectiveness and budget impact. [99] As for many schemes, increasing the number of patients for whom negotiations are held is thought to increase bargaining power and decrease prices, since companies are then less likely to deflect a low offer due to fears of losing market share. [84;190;196] New Zealand implemented a pharmaceutical management agency to have a single negotiator for the entire population,

but not all countries are so organized. [78] For example, in the US, Medicare is not allowed to negotiate for drug prices[119;197]; if a drug is approved and prescribed, Medicare has to cover it, so its bargaining power is limited. [119] Other players in the highly fragmented US market are small, thus reducing everyone's negotiating power. [98]

An important downside to negotiating lower prices on a health care provider level or national level is that this might indirectly increase prices for other providers and countries, who have smaller health care budgets, which results in a weaker position to negotiate. [1;145]

4.2.5.2.7 Transnational licensing and pricing frameworks

To increase access to drugs that are on patent and expensive, but necessary in LMIC, these countries' authorities can choose to issue compulsory licences as allowed by the World Health Organization (WHO). [79;198] This means that the authorities recognize the drug patents, but are allowed to have local generic manufacturers produce the same drugs, without fearing claims of patent infringement, or they can import the drug from another generic manufacturer. [7;56;79;81] This reduces the costs of a new drug dramatically, [171] though other options like international procurement seem to offer a better discount. Unfortunately, this approach is also administratively cumbersome, since in general, it applies to one drug at a time, [56] and could result in other innovators withdrawing their drug from the market. [199] However, compulsory licensing can be used successfully as part of a strategy to reduce prices offered by the originator. [56;79;81;199]

International procurement is based on collective price negotiations between an innovative company and a union of LMIC. This approach leads to lower prices and more accessibility than compulsory licensing. [81] Lower prices can be achieved through voluntary out-licensing, [73] wherein the originator allows a generic manufacturer to produce the drug at reduced costs in exchange for a royalty. One example is the out-licensing of Harvoni® (containing sofosbuvir and ledipasvir), which Gilead Sciences gave to an Indian manufacturer to produce for 91 LMIC, against a royalty of 7%. [81]

High-income countries can also benefit from forming a union to increase bargaining power. For example, The Netherlands and Belgium recently signed an agreement to negotiate process for orphan drugs as a block. Several EU-countries have followed this example and joined the

agreement, and some pharmaceutical companies have indicated their willingness to cooperate. [150]

A more utopian option to regulate drug pricing is the proposal of a tiered pricing framework. [79;80] This would put a global public body, such as the WHO, in charge of regulating prices of patented drugs worldwide. It would set prices for countries based on income, disease burden and possibly the rates of out-of-pocket payments. This would differentiate prices for rich and poor countries, and achieve fairer drug prices for consumers on a global scale.

Difficulties in the execution of this plan are to obtain international consent on the rules and calculations, and prevent leakage into other markets through parallel trade. Also, patent rights would have to be equally respected and be applicable for the same terms, since it would be a void policy if patent-infringing cheaper generic drugs were available. [80] Analyses of existing systems with a comparable design in LMIC have shown increased profits for pharmaceutical companies and an increased availability of medical innovations. [80]

A comparable idea is that a global fund financed by governments would reward companies, based on a share of the contribution they make to global health with all their products. In exchange, those companies would have to manufacture a drug at the lowest feasible price. [1;200] Unfortunately, this approach would be extremely difficult to implement and measure.

4.2.6 Discussion

Taking into account the complex and interconnected dynamics of drug life cycles, pricing and intervention policies, several issues are worth mentioning and questioning. Recurring arguments in the drug pricing debate are now presented for discussion.

4.2.6.1 R&D through merger and acquisition

Companies are demonstrating a shift from in-house R&D to cheaper merger and acquisition-based development to fill the pipeline. [41;42;104;201] This shift is due to the reduced efficiency of in-house research, [86] and transfers the risk of failing research from big companies to small startups. In this new format, small startups go bankrupt when the research does not yield a profitable product, so that large companies don't have to suffer the losses. In the case of a successful start-up, larger companies simply buy the licence or the entire company.

In the end risk is shifted from big manufacturers to investors and governments who have supported those startups. Although this business strategy is understandable in economic terms, it provokes perverse effects in the pharmaceutical marketplace that require new forms of regulation.

Valeant, [45;114;202] Turing and Amedra [197] are examples of companies that are mostly focused on buying profitable products, instead of performing in-house R&D. Another example is Gilead Sciences, which bought sofosbuvir (Sovaldi®) from Pharmasset, and marketed the drug at double the cost that Pharmasset had intended to charge. [203] We seriously doubt the long-term viability of investor-centered business strategy. [147;204]

Some generics are only produced by a couple of companies, so it is possible to buy all the drug rights and drastically raise prices. [16;140;184] This practice allows for a monopoly until competitors succeed in starting production as well. A major delay in this is the processing time for generic approvals, which is approximately ten months. [140] Many of the recent overnight increases in drug prices have been caused by this tactic that closely resembles a hedge fund strategy, rather than a pharmaceutical one. [16;200] This strategy was used for Daraprim®, [15;16] albendazole, [197] Aloquin, [20] doxycycline and thalidomide. [140]

4.2.6.2 Patent law revision and stimulating public-private partnerships

The patent system is often blamed for high prices, because it limits competition and helps create monopolies. There certainly is an urgent need to revise the patent system in order to stimulate true innovation and prevent surrogate innovation as well as abuse of the patent system by a pharmaceutical industry that increasingly seeks to exploit the re-patenting loophole. [205] Unfortunately, removing the patent system completely would significantly reduce companies' incentive to invest in research, so other solutions are required. [41] An alternative trajectory might be to have all drug research paid for by public money, in the same way that outcomes – prescription drugs – are currently paid for by public money. Academic institutions already perform a significant portion of new drug development, but currently lack the funds, capacity and incentives to develop a drug completely without support. [41;206] Partnerships between companies, governments, research and charity organizations, referred to as public-private financing partnerships, with

agreements on the drug availability and product price, seem most promising. [193;194] This approach is what accelerated the development of the malaria vaccine, Mosquirix®.

New initiatives to help reduce patents' limiting accessibility and affordability effects are currently being developed. [32] GSK announced a graduated approach, stating that it would not defend patents in the poorest countries, and transfer licences to generic manufacturers in LMIC using the World Bank classification. [207-209] This approach would increase access to new drugs without limiting profitability in high-income countries.

Public-private partnerships could also stimulate commercially unattractive but essential therapeutic innovations in high-income countries. Exemplary is the clear societal need for new antibiotics, given the increasing prevalence of multi-drug-resistant bacteria. The possible gains for the industry are too low to make innovation economically viable. [200] The population base for a new antibiotic is small, because it would be the option of last resort. In addition, the length of antibiotic treatment is short, making the volume of the market very small. [210] Extremely high prices could compensate for this situation, but it is doubtful whether society would be willing to pay. Due to low incentives in the current market, governments are tempted to impose regulation to make antibiotic development more attractive, [206;210;211] as is in place for orphan drugs. We recommend that this policy also includes strict drug pricing conditions, and that it is accompanied by antimicrobial stewardship to prevent the overuse and misuse of the new antibiotic drug.

4.2.6.3 Me-too drugs, R&D resources

Me-too drugs or follow-on drugs are drugs with minor chemical variations relative to a drug already on the market within a given therapeutic class. [173] These drugs are highly controversial since they often cost roughly the same as the first-in-class drugs, but offer few relevant therapeutic improvements. [68;69] Me-too drugs are seen as an ineffective consumption of R&D resources and diminishing incentives for innovation. [24]

Another side of the argument is that me-too drugs increase choice, and make treatment available for certain groups, or help match a drug's pharmacokinetics, effectiveness or side effects in specific

populations. [24;68] Me-too drugs are also a consequence of R&D races between multiple companies developing drugs for the same indication, so they are inherent to a competitive system. [173;212] Given that a first-in-class drug has a significant advantage in market share and costs, this competition forces companies to speed up innovation. [69]

Studies suggest that the price of me-too drugs only falls after the third introduction. [93;173] Where the first movers compete over quality only, the drugs introduced later compete over price to overcome the disadvantage of being a new drug with no clinical record. [212] This holds for Gilead Sciences' hepatitis C drugs as well, where MSD is competing with Zepatier® with a price tag that is 40% lower than the first-in-class. [115;213]

4.2.6.4 Lack of systematic research into policy effectiveness

Various policies related to drug prices with roughly the same aim have been introduced around the world, but comparing the effectiveness across policies appears difficult. [41] Policies often lack a scientific basis, and an evaluation after the policy is introduced is not always performed, let alone in a standardized way. [41;84;165] Insights on the long-term policy effects of for instance value-based (QALY-guided) pricing are particularly scarce, [165] which causes some countries to pay unnecessarily high prices for medicines. [35;163] Even regulatory guidelines are implemented without evaluation, [41] which causes some to ask for lowered regulation to speed up innovation[66] and others to ask for more regulation for additional safety. Unfortunately, both groups lack the evidence to back their case. Given the current trends in big data analytics, better monitoring of the effects of drugs with respect to health outcomes and the effects of policy on pricing should be possible. [89] Health Technology Assessment methodologies (HTA) could be used to strengthen evaluation of policy measures. Furthermore, the option of involving regulatory gatekeepers of safety and effectiveness like the FDA and the EMA in drug pricing policies should be considered more seriously. Like drugs, drug policy must be evidence-based. [110]

4.2.6.5 Justification of drug prices

Life expectancy has gone up by 30 years in just a century in high-income countries. [77] It can be stated that innovation is always expensive, as seen in other new technologies, and once the price of innovation has been paid, generics make innovative treatments widely available. Revolutions in biotechnology, nanotechnology, biophysics and genome sequencing

have moved precision medicine from the bench to the bedside, [53;89;214] yielding a burst of innovation in medicine. [30;42;75;160] This helps to explain the large number of new and expensive targeted therapies that have come to the market recently. Cancer mortality has been reduced by 20% in the last two decades[77] through the introduction of new drugs, but that is also due to screening, prevention, vaccination and surgical improvements. [58]

Steep pricing strategies are historically an oncological phenomenon, [91;214] and recently prices are rising to even higher levels. [94] This is peculiar, since some oncological drugs offer only small benefits. [5] In addition to this low effect, oncological drugs are frequently priced so high that they are not cost-effective. [215] What makes the situation more financially challenging is that cancer drugs are often more effective in combinations, [75] thus making therapy cost the sum of the drugs. [98] Debates about this situation have usually ended in the consensus that patients cannot be denied new cancer treatments simply because of costs, [214] so eventually most cancer drugs have been reimbursed despite the high costs. But this is no longer sustainable given the steadily growing patient populations with cancer and stagnating health budgets.

As more treatments become possible, expectations rise. [58] Diseases that were a death sentence decades ago are now treatable, and leaving them untreated is not acceptable. With increasing possibilities come increased demands and increasing costs. [216]

These innovations have allowed for more personalized diagnosis and cures, leading to precision or personal medicine, aimed at highly stratified patient populations. This gives better outcomes, but given the often smaller populations that qualify for a treatment (and thus smaller volumes of sales) prices are high to generate adequate revenue. [214] This shift from blockbuster to niche buster will make many untreatable diseases treatable, but costs for the healthcare sector will continue to rise. [217]

4.2.6.6 Patient-assistance programs and list prices

Companies that raise prices often defend their actions by stating that patients who cannot afford the drugs are offered assistance in the form of patient assistance programs programmes in Western countries. [156] These programs allow patients to apply for the drugs at reduced or no

cost, if they are uninsured and live below a certain income level. [114] The income level is set so that many patients on normal wages don't qualify, so that drug prices can result in catastrophic spending. [114] Furthermore, patient-assistance programs increase the workload for general practitioners' assistants, since they often require many forms. [140] The costs to the healthcare system are still unnecessarily large, and are shifted from patients to insurance companies. [141;154;218] These programs basically allow companies to sponsor the purchase of their own drugs. [147]

Another argument for defending high prices is that pharmaceutical companies do not actually charge list prices for drugs. Hospitals and insurers often negotiate discounts on drug list prices, [142;156;219] and sometimes up to 50%. [141] However, drug prices worldwide remain at unrealistically high levels, since examples of annually raising prices (or prices rising after selling licences) to more than double the original price are plentiful. [114;220]

In some cases, companies provide free drugs for a small market simply because they feel they are morally obliged to do so. These are usually small companies who do not want to invest in regulatory approval procedures. [70] Larger companies see it as a moral obligation to maximize sales volumes and profits, reach patients and secure future investments in research, continue innovation, [94] and pay their stock holders.

Trying to defend high spending on pharmaceuticals, companies often point out that in the US only $300 billion is spent annually on prescription medicines compared to the $1 trillion that is spent on hospital care. [2;184] Therefore, compared to total healthcare spending, US pharmaceutical costs are just 10% of the total health care budget, which is roughly the same throughout Europe. [184] This number has remained constant for over 50 years. [77;221] Thus, all other healthcare costs, such as hospitals and general practitioners' costs, have risen at the same rate as pharmaceutical costs. [39] It is simply easier for governments to reduce drug budgets than to reduce hospital staff wages or restructure inefficient healthcare infrastructures. [2] However, this still does not justify high drug prices. Just because other parts of the system are not functioning as efficiently as possible, does not mean pharmaceutical companies should be increasing prices at the same rate.

4.2.6.7 Methodological notes

This paper used both scientific publications and newspaper articles from several selected newspapers. The selection of those sources allowed for a comprehensive view on the subject of drug pricing. Clear differences in focus between the newspaper articles and scientific publications were observed.

The newspaper articles tended to focus on public opinion in one country based on one event, with examples and personal stories about the impact of high drug prices on patients' lives. A thorough analysis on causes and policies of drug pricing was often missing, and a rather mono-dimensional conclusion was frequently offered that pharmaceutical companies choose to keep the prices high for their own exorbitant profits.

The journal articles offered more policy-focussed analyses of the causes of and solutions to high drug pricing. They more frequently covered in-depth discussions of single policy measures per country and their proposed or measured effect, but often left the patient perspective out.

In conclusion, the combination of both peer-reviewed scientific papers and newspaper articles allowed for a significantly contextualized evidence-based conclusion. The mixed-method approach yielded deeper insights into the problem area of high drug pricing and sustainable drug markets than one of the two individual methods would have been able to achieve alone.

4.2.7 Conclusion

The current rise in drug prices worldwide is making healthcare unaffordable even in high-income countries. Apart from historic changes in the drug life cycle dynamics, price-volume proportions, and a transition from "one-size-fits-all" to more stratified precision medicine approaches, this problem is due to patent-induced monopoly positions, unintended consequences of drug and reimbursement policies and competitive market failure. This situation threatens to disturb the fragile compromise between the basic human right for affordable access to healthcare and the utilitarian protection of inventions to incentivize innovation. The current pricing spiral will only stop through well-designed regulatory interventions and measures around drug pricing on a national and transnational levels. Access to medicines needs to be

central to any policy intervention discussion — something that can be overlooked when governments are arguing for lower health care costs to reduce spending, while the pharmaceutical industry is repeating the argument for competitive pricing to reimburse their R&D costs.

Many options to regulate the pharmaceutical market have been tried, some with better success than others. It is clear that reference pricing – both internal and external – and incentivizing physicians and pharmacists are the fastest, most effective and most reviewed options. Transnational cooperation, for example through the European Union, African Union, World Bank or WHO, would help reduce drug prices with increased bargaining power, and could potentially reduce administrative costs. Apart from creating collective negotiating power transnational cooperation would also stimulate exchanging trial data, sharing patient records and improving evaluation methods. A global framework for cooperation among drug regulatory authorities (e.g. FDA and EMA) would increase those benefits even more, and could be further amplified by reinforcing the existing WHO framework that already helps to reduce drug prices by means of an essential medicines list, which facilitates compulsory licensing.

Reduced healthcare spending is thought to reduce incentives for innovation, but given the current double-digit profit margins, industrial incomes could be lower without harming the industry's outlooks. Public-private partnerships, in which charity funds are used to sponsor research in exchange for lower prices, could significantly help direct spending decisions on research away from primarily financial motives towards what is best for society.

Value-based pricing is a promising but also risky option that is already being used by some countries to reduce costs. The (inter-)national public debates about how much a QALY should cost and the regulatory and policy debates about whether and how to continue with the QALY appraisal tools still have to reach a conclusion. Until consensus is reached, drug companies will continue to strategically use the QALY and/or ICER thresholds to boost their profits. Governments should take this into account while continuing to deploy sophisticated ICER-based compensation models in the era of precision medicine.

In conclusion, the recent rise in drug prices is caused by uncontrolled market dynamics, changes in life-cycle dynamics and

unanticipated policy side-effects. There is a wide range of policy tools to reduce drug prices available through various mechanisms. The most effective options are reference pricing and incentivizing physicians and pharmacists, whereas for the long term value-based and outcome-based pricing next to public-private partnerships are most promising developments. The challenge is, of course, how to strike a balance between rewarding investments in innovation, achieving reasonable drug pricing for governments and securing equitable access to medicines.

4.2.7.1 Acknowledgements

The authors would like to thank Cassandra Nemzoff, Nathalie Kuijpers and Julia Challinor for their English manuscript correction services. In addition, we would like to thank Frank-Jan van Lunteren for his graphics.

4.2.7.2 Funding

This study was performed in the context of regular research of the Division of Pharmacoepidemiology and Clinical Pharmacology (Utrecht University), employing authors TvdG and TP, and of the Institute for Medical Technology Assessment, Department of Health Policy & Management, Erasmus University, Rotterdam, employing CU. The university had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

4.2.7.3 Conflict of interest

None declared.

4.2.7.4 Author Contributions

TvdG and TP designed the research, TvdG and TP conducted the research and analyzed data. TvdG and TP wrote the paper. TvdG, TP and CU edited the manuscript. TvdG and TP had primary responsibility for final content. All authors have read and approved the final manuscript.

4.2.8 References

1. Parker-Lue S, Santoro M, Koski G. The ethics and economics of pharmaceutical pricing. Annu Rev Pharmacol Toxicol 2015;55:191-206.
2. Andrew Ward. Pharmaceuticals: Value over volume. Financial Times 2015 Sep 24.
3. Pollack A. Senators Condemn Price Rises for Drugs. The New York Times 2015 Dec 10.
4. Kushnick HL. Pricing Cancer Drugs: When Does Pricing Become Profiteering? AMA J Ethics 2015 Aug;17(8):750-3.

5. Barlas S. Are specialty drug prices destroying insurers and hurting consumers?: a number of efforts are under way to reduce price pressure. P T 2014 Aug;39(8):563-6.
6. Experts in Chronic Myeloid Leukemia. The price of drugs for chronic myeloid leukemia (CML) is a reflection of the unsustainable prices of cancer drugs: from the perspective of a large group of CML experts. Blood 2013 May 30;(1528-0020 (Electronic)).
7. Radhakrishnan P. Commentary: Making middle income countries pay full price for drugs is a big mistake. BMJ 2015;351:h3757.
8. Wilensky G. A New Focus on Prescription Drug Spending. JAMA 2015 Aug 4;314(5):440-1.
9. Committee on Finance USS. The price of Sovaldi and its impact on the U.S. healt care system. Washington: U.S. Government Publishing Office; 2015 Dec.
10. Neumann PJ, Cohen JT. Measuring the Value of Prescription Drugs. N Engl J Med 2015 Nov 18.
11. Boseley S. Hepatitis C drug delayed by NHS due to high cost. The Guardian 2015 Jan 16.
12. Hawlik K. Access to High-priced Medicines in Hospital Settings in Europe: A Study in Four European Countries. Amsterdam: Health Action International; 2016 Feb 22.
13. Barlas S. States try to control medicaid pharmaceutical costs: numerous, diverse cost pressures force myriad reform efforts. P T 2015 Apr;40(4):260-2.
14. The Editorial Board. Runaway Drug Prices. The New York Times 2015 May 5.
15. Farrell S. Clinton's promise on treatment prices hits pharma shares. The Guardian 2015 Sep 22.
16. Pollack A. An Executive's Arrest Gives Drug Makers Cover. The New York Times 2015 Dec 18.
17. Pollack A. Price Increase Rescinded for a Tuberculosis Drug. The New York Times 2015 Sep 22.
18. Walker J, Winslow R, Steele A. Mylan to Offer Generic EpiPen. Wall Street Journal 2016 Aug 30.
19. Rockoff JD. The Year in Review: Corporate News: Drug Prices Draw New Fire. Wall Street Journal 2016 Dec 21.
20. Neate R. US drug company hiked price of acne cream by 3,900% in less than 18 months. The Guardian 2016 Sep 21.
21. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. N Engl J Med 2017 Mar 17.
22. Kolata G. Cholesterol-Slashing Drug Can Protect High-Risk Heart Patients, Study Finds. The New York Times 2017 Mar 17.

23. Wirtz VJ, Hogerzeil HV, Gray AL, Bigdeli M, de Joncheere CP, Ewen MA, et al. Essential medicines for universal health coverage. The Lancet 2016 Nov 7;389(10067):403-76.
24. Brooks E, Geyer R. Can a medical need clause help manage the growing costs of prescription drugs in the EU? Health Econ Policy Law 2015 Sep 3;1-14.
25. Yoo KB, Lee SG, Park S, Kim TH, Ahn J, Cho MH, et al. Effects of drug price reduction and prescribing restrictions on expenditures and utilisation of antihypertensive drugs in Korea. BMJ Open 2015;5(7):e006940.
26. Silverman E. Tools for Taking the Measure Of Cancer Drugs. Manag Care 2015 Jul;24(7):7.
27. Loftus P. Gilead Knew Hepatitis Drug Price Was High, Senate Says. The Wall Street Journal 2015 Dec 1.
28. AJMC Staff. Kadcyla and Opdivo Too Expensive for NICE. The American Journal of Managed Care . 16-12-5015.
29. Boseley S. Breast cancer drug rejected for NHS use on cost-benefit grounds. The Guardian 2016 Dec 29.
30. Rankin J. Cancer breakthroughs trigger big pharma interest in drugs and deals. The Guardian 2015 Jun 1.
31. Røe OD. The high cost of new cancer therapies - a challenge of inequality for all countries. JAMA Oncology 2016 Dec 29.
32. Ploumen L, Schippers E. Better life through medicine - Let's leave no one behind. The Lancet 2017 Jan 28;389(10067):339-41.
33. Levy M, Rizansky NA. The pricing of breakthrough drugs: theory and policy implications. PLoS One 2014;9(11):e113894.
34. Han E, Chae SM, Kim NS, Park S. Effects of pharmaceutical cost containment policies on doctors' prescribing behavior: Focus on antibiotics. Health Policy 2015 Sep;119(9):1245-54.
35. Miziara NM, Coutinho DR. Problems in the regulatory policy of the drug market. Rev Saude Publica 2015;49:35.
36. Acosta A, Ciapponi A, Aaserud M, Vietto V, Austvoll-Dahlgren A, Kosters JP, et al. Pharmaceutical policies: effects of reference pricing, other pricing, and purchasing policies. Cochrane Database Syst Rev 2014;10:CD005979.
37. Khatib R, McKee M, Shannon H, Chow C, Rangarajan S, Teo K, et al. Availability and affordability of cardiovascular disease medicines and their effect on use in high-income, middle-income, and low-income countries: an analysis of the PURE study data. Lancet 2015 Oct 20.
38. Kohler JC, Mitsakakis N, Saadat F, Byng D, Martinez MG. Does Pharmaceutical Pricing Transparency Matter? Examining Brazil's Public Procurement System. Global Health 2015;11:34.
39. Adams KT. Is research to blame for high cost of cancer drugs? Manag Care 2014 Jun;23(6):46-8.

40. Reinhardt U. Probing Our Moral Values in Health Care: The Pricing of Specialty Drugs. JAMA 2015 Sep 8;314(10):981-2.
41. Moors EH, Cohen AF, Schellekens H. Towards a sustainable system of drug development. Drug Discov Today 2014 Nov;19(11):1711-20.
42. Andrew Ward. Healthcare: Counting the cost of cancer - FT. Financial Times 2015 Jan 15.
43. Andrew Ward, David Crow. AstraZeneca chief warns UK 'falling behind' in cancer care. Financial Times 2015 Jun 2.
44. David Crow. New heart drugs set to add billions to US healthcare bill. Financial Times 2015 Jun 10.
45. Jonathan Ford. Valeant and other raiders of pharma rust belt are a threat to R&D. Financial Times 2015 Nov 1.
46. Shrank WH, Barlow JF, Brennan TA. New Therapies in the Treatment of High Cholesterol: An Argument to Return to Goal-Based Lipid Guidelines. JAMA 2015 Oct 13;314(14):1443-4.
47. Goodman J, Walsh V. The Story of Taxol: Nature and politics in the pursuit of an anti-cancer drug. Cambridge: 2001.
48. Pieters T. Marketing medicines through randomised controlled trials: the case of interferon. BMJ 1998 Oct 31;317(1231).
49. Bae G, Park C, Lee H, Han E, Kim DS, Jang S. Effective policy initiatives to constrain lipid-lowering drug expenditure growth in South Korea. BMC Health Serv Res 2014;14:100.
50. Cho MH, Yoo KB, Lee HY, Lee KS, Kwon JA, Han KT, et al. The effect of new drug pricing systems and new reimbursement guidelines on pharmaceutical expenditures and prescribing behavior among hypertensive patients in Korea. Health Policy 2015 May;119(5):604-11.
51. Ben-Aharon O, Shavit O, Magnezi R. Does drug price-regulation affect healthcare expenditures? Eur J Health Econ 2016 Sep 30.
52. Abramowitz PW, Cobaugh DJ. The costs of prescription drugs in the United States: Pharmacists' voices must be heard. Am J Health Syst Pharm 2016 Apr 15;73(8):515-6.
53. Ramsey SD. How state and federal policies as well as advances in genome science contribute to the high cost of cancer drugs. Health Aff (Millwood) 2015 Apr;34(4):571-5.
54. Godman B, Wettermark B, van WM, Fraeyman J, Alvarez-Madrazo S, Berg C, et al. Multiple policies to enhance prescribing efficiency for established medicines in Europe with a particular focus on demand-side measures: findings and future implications. Front Pharmacol 2014;5:106.
55. Rudisill C, Vandoros S, Antoun JG. Pharmaceutical policy reform in the Russian Federation. J Health Polit Policy Law 2014 Jun;39(3):691-705.

56. Sruamsiri R, Ross-Degnan D, Lu CY, Chaiyakunapruk N, Wagner AK. Policies and programs to facilitate access to targeted cancer therapies in Thailand. PLoS One 2015;10(3):e0119945.
57. John Gapper. The unhealthily high price of cancer drugs. Financial Times 2015 Jun 3.
58. Schnipper LE, Davidson NE, Wollins DS, Tyne C, Blayney DW, Blum D, et al. American Society of Clinical Oncology Statement: A Conceptual Framework to Assess the Value of Cancer Treatment Options. Journal of Clinical Oncology 2015 Aug 10;(1527-7755 (Electronic)).
59. Kaiser U, Mendez SJ, Ronde T, Ullrich H. Regulation of pharmaceutical prices: evidence from a reference price reform in Denmark. J Health Econ 2014 Jul;36:174-87.
60. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 2009 Jul 21;6(7):e1000097.
61. Barksdale HC, Harris CE. Portfolio analysis and the product life cycle. Long Range Planning 1982;74-83.
62. Pieters T. Interferon: The science and selling of a miracle drug. London: Routledge; 2005.
63. Waber RL, Shiv B, Carmon Z, Ariely D. Commercial features of placebo and therapeutic efficacy. JAMA 2008 Mar 5;299(9):1016-7.
64. Voet MA. The generic challange: understanding patents, FDA and pharmaceutical life-cycle management. Fourth edition ed. Boca Raton, Florida: BrownWalker Press; 2014.
65. Falit BP, Singh SC, Brennan TA. Biosimilar competition in the United States: statutory incentives, payers, and pharmacy benefit managers. Health Aff (Millwood) 2015 Feb;34(2):294-301.
66. Berndt ER, Nass D, Kleinrock M, Aitken M. Decline in economic returns from new drugs raises questions about sustaining innovations. Health Aff (Millwood) 2015 Feb;34(2):245-52.
67. Kakkar AK. Patent cliff mitigation strategies: giving new life to blockbusters. Expert Opin Ther Pat 2015 Sep 15;1-4.
68. Nutt DJ, Attridge J. CNS drug development in Europe--past progress and future challenges. Neurobiol Dis 2014 Jan;61:6-20.
69. Andrade LF, Sermet C, Pichetti S. Entry time effects and follow-on drug competition. Eur J Health Econ 2014 Dec 12.
70. Tavernise S. Patients Fear Spike in Price of Old Drugs. The New York Times 2015 Dec 23.
71. Feldschreiber P, Breckenridge A. After thalidomide - do we have the right balance between public health and intellectual property. Rev Recent Clin Trials 2015;10(1):15-8.

72. Ciociola AA, Cohen LB, Kulkarni P. How drugs are developed and approved by the FDA: current process and future directions. Am J Gastroenterol 2014 May;109(5):620-3.
73. Berndt ER, Cockburn IM. The hidden cost of low prices: limited access to new drugs in India. Health Aff (Millwood) 2014 Sep;33(9):1567-75.
74. Lachmann PJ. A more radical solution. Rev Recent Clin Trials 2015;10(1):25-7.
75. Robert C, Karaszewska BF, Schachter JF, Rutkowski PF, Mackiewicz AF, Stroiakovski DF, et al. Improved overall survival in melanoma with combined dabrafenib and trametinib. N Engl J Med 2015 Jan 1;(372):30-9.
76. Jekunen A. Decision-making in product portfolios of pharmaceutical research and development--managing streams of innovation in highly regulated markets. Drug Des Devel Ther 2014;8:2009-16.
77. Ingram R. A Not-So-Transparent Attempt to Cap Drug Prices. The Wall Street Journal 2015 Jul 19.
78. Vogler S, Kilpatrick K, Babar ZU. Analysis of Medicine Prices in New Zealand and 16 European Countries. Value Health 2015 Jun;18(4):484-92.
79. Ooms G, Forman L, Williams OD, Hill PS. Could international compulsory licensing reconcile tiered pricing of pharmaceuticals with the right to health? BMC Int Health Hum Rights 2014;14:37.
80. Williams OD, Ooms G, Hill PS. Cautionary Notes on a Global Tiered Pricing Framework for Medicines. Am J Public Health 2015 Jul;105(7):1290-3.
81. Beall RF, Kuhn R, Attaran A. Compulsory licensing often did not produce lower prices for antiretrovirals compared to international procurement. Health Aff (Millwood) 2015 Mar;34(3):493-501.
82. Miller G. Is pharma running out of brainy ideas? Science 2010 Jul 30;329(5991):502-4.
83. Messner DA, Towse A, Mohr P, Garau M. The future of comparative effectiveness and relative efficacy of drugs: an international perspective. J Comp Eff Res 2015;4(4):419-27.
84. Srivastava D, McGuire A. Analysis of prices paid by low-income countries - how price sensitive is government demand for medicines? BMC Public Health 2014;14:767.
85. Bagel J. Biosimilars. J Drugs Dermatol 2014 Jul;13(7):788-90.
86. Elkind P, Reingold J. Inside Pfizer's palace coup. Fortune 2011 Aug 15.
87. Pieters T, Snellers S. Managing double binds in the pharmaceutical prescription market: the case of Halcion. In: Gaudillière J, Hess V, editors. Ways of regulating drugs in the 19th and 20th centuries.London: 2012. p. 270-86.

88. Lincker H, Ziogas C, Carr M, Porta N, Eichler HG. Regulatory watch: Where do new medicines originate from in the EU? Nat Rev Drug Discov 2014 Feb;13(2):92-3.
89. Tormay P. Big Data in Pharmaceutical R&D: Creating a Sustainable R&D Engine. Pharmaceut Med 2015;29(2):87-92.
90. Levaggi R. Pricing schemes for new drugs: a welfare analysis. Soc Sci Med 2014 Feb;102:69-73.
91. David Crow, Andrew Ward. Healthcare: The race to cure rising drug costs. Financial Times 2015 Feb 10.
92. Puig-Junoy J, Lopez-Valcarcel BG. Launch prices for new pharmaceuticals in the heavily regulated and subsidized Spanish market, 1995-2007. Health Policy 2014 Jun;116(2-3):170-81.
93. Shajarizadeh A, Hollis A. Price-cap Regulation, Uncertainty and the Price Evolution of New Pharmaceuticals. Health Econ 2015 Aug;24(8):966-77.
94. Rockoff J. How Pfizer Set the Cost of Its New Drug at $9,850 a Month. The Wall Street Journal 2015 Dec 9.
95. Hu S, Zhang Y, He J, Du L, Xu M, Xie C, et al. A Case Study of Pharmaceutical Pricing in China: Setting the Price for Off-Patent Originators. Appl Health Econ Health Policy 2015 Aug;13 Suppl 1:S13-S20.
96. Kumar R, Hassali MA, Saleem F, Alrasheedy AA, Kaur N, Wong ZY, et al. Knowledge and perceptions of physicians from private medical centres towards generic medicines: a nationwide survey from Malaysia. J Pharm Policy Pract 2015;8(1):11.
97. Hennessy D, Sanmartin C, Ronksley P, Weaver R, Campbell D, Manns B, et al. Out-of-pocket spending on drugs and pharmaceutical products and cost-related prescription non-adherence among Canadians with chronic disease. Health Rep 2016 Jun 15;27(6):3-8.
98. Whalen J. Doctors Object to High Cancer-Drug Prices. The Wall Street Journal 2015 Jul 23.
99. Wouters OJ, Kanavos PG. Transitioning to a national health system in Cyprus: a stakeholder analysis of pharmaceutical policy reform. Bull World Health Organ 2015 Sep 1;93(9):606-13.
100. Pollack A. Drug Makers Sidestep Barriers on Pricing. The New York Times 2015 Oct 20.
101. Duso T, Herr A, Suppliet M. The welfare impact of parallel imports: a structural approach applied to the German market for oral anti-diabetics. Health Econ 2014 Sep;23(9):1036-57.
102. Amy Kazmin. Pharma combats diversion of cheap drugs. Financial Times 2015 Apr 12.

103. Sheingold S, Nguyen NX. Impacts of generic competition and benefit management practices on spending for prescription drugs: evidence from Medicare's Part D benefit. Medicare Medicaid Res Rev 2014;4(1).
104. Walters J. Price-hiking drug companies put 'lives in the balance', US lawmakers hear. The Guardian 2015 Dec 10.
105. Emanuel E. I Am Paying for Your Expensive Medicine. The New York Times 2015 Nov 8.
106. McGee S. Investigating the mystery of soaring generic medication prices. The Guardian 2015 Oct 11.
107. Loftus P. How Much Should Cancer Drugs Cost? The Wall Street Journal 2015 Jun 18.
108. Tallapragada NP. Off-patent drugs at brand-name prices: a puzzle for policymakers. J Law Biosci 2016 Apr;3(1):238-47.
109. Editorial. Paying a high price for cancer drugs. Lancet 2015 Aug 1;386(9992):404.
110. Mailankody S, Prasad V. Implications of Proposed Medicare Reforms to Counteract High Cancer Drug Prices. JAMA 2016 Jul 19;316(3):271-2.
111. Boseley S. Cancer drug prices must come down, say leading research institutes. The Guardian 2017 Feb 9.
112. Pollack A. Drug Companies Increasingly Pushed to Explain High Prices. The New York Times 2015 Jul 23.
113. Sharma A, Rorden L, Ewen M, Laing R. Evaluating availability and price of essential medicines in Boston area (Massachusetts, USA) using WHO/HAI methodology. J Pharm Policy Pract 2016;9:12.
114. David Crow. Valeant's business model faces tough questions. Financial Times 2015 Oct 8.
115. Eichler HG, Hurts H, Broich K, Rasi G. Drug Regulation and Pricing--Can Regulators Influence Affordability? N Engl J Med 2016 May 12;374(19):1807-9.
116. Mulinari S. Divergence and convergence of commercial and scientific priorities in drug development: The case of Zelmid, the first SSRI antidepressant. Soc Sci Med 2015 Aug;138:217-24.
117. Tadena N. Doctor's Proposed Ban Of Drug Ads Goes After Top Magazine Ad Category. The Wall Street Journal 2015 Nov 18.
118. Reinke T. Spending on Compounded Drugs Goes Sky High. Manag Care 2015 May;24(5):18-9.
119. Whalen J. Why the U.S. Pays More Than Other Countries For Drugs. The Wall Street Journal 2015 Dec 1.
120. European Commission. Legal framework governing medicinal products for human use in the EU. EC 2016 March 28Available from: URL: http://ec.europa.eu/health/human-use/legal-framework/index_en.html

121. Boseley S. UK NHS cancer patients denied drugs due to inflated prices – experts. The Guardian 2015 Sep 23.
122. Hill KP, Ross JS FAU - Egilman D, Egilman DS FAU - Krumholz H, Krumholz HM. The ADVANTAGE seeding trial: a review of internal documents. Annals of Internal Medicine 2008 Aug 19;149(1539-3704 (Electronic)):251-8.
123. Katz KA. Time to nip "seeding trials" in the bud. Arch Dermatol 2008 Mar 1;144(3)(1538-3652 (Electronic)):403-4.
124. Saastamoinen LK, Verho J. Register-based indicators for potentially inappropriate medication in high-cost patients with excessive polypharmacy. Pharmacoepidemiol Drug Saf 2015 Jun;24(6):610-8.
125. Perlis RH, Perlis CS. Physician Payments from Industry Are Associated with Greater Medicare Part D Prescribing Costs. PLoS One 2016;11(5):e0155474.
126. van Harten WH, Wind A, de PP, Saghatchian M, Oberst S. Actual costs of cancer drugs in 15 European countries. Lancet Oncol 2016 Jan;17(1):18-20.
127. Walton SM, Rash C, Lambert BL, Galanter WL. A case study in generic drug use: should there be risk adjustment in incentive payments for the use of generic medications? J Manag Care Spec Pharm 2014 Nov;20(11):1093-9.
128. Toverud EL, Hartmann K, Hakonsen H. A Systematic Review of Physicians' and Pharmacists' Perspectives on Generic Drug Use: What are the Global Challenges? Appl Health Econ Health Policy 2015 Aug;13 Suppl 1:S35-S45.
129. Yokoi M, Tashiro T. Prescription, Dispensation, and Generic Medicine Replacement Ratios: Influence on Japanese Medicine Costs. Glob J Health Sci 2015;8(1):45590.
130. Dylst P, Vulto A, Simoens S. Does increased use of generic medicines by elders in Belgium help to contain escalating health care budgets? J Aging Soc Policy 2014;26(3):266-80.
131. Kalo Z, Holtorf AP, Alfonso-Cristancho R, Shen J, Agh T, Inotai A, et al. Need for multicriteria evaluation of generic drug policies. Value Health 2015 Mar;18(2):346-51.
132. Nardi EP, Ferraz MB, Pinheiro GR, Kowalski SC, Sato EI. Perceptions of the population regarding generic drugs in Brazil: a nationwide survey. BMC Public Health 2015;15:117.
133. Drozdowska A, Hermanowski T. Predictors of generic substitution: The role of psychological, sociodemographic, and contextual factors. Res Social Adm Pharm 2015 Mar 27.
134. Vij R. Pharma Industry Merger And Acquisition Analysis 1995 To 2015. www revenuesandprofits com 2016 [cited 2016 Jun 2];Available from: URL: www.revenuesandprofits.com

135. Howie L, Hirsch B, Abernethy A. A comparison of FDA and EMA drug approval: implications for drug development and cost of care. Oncology (Williston Park) 2013 Dec 5;(0890-9091 (Print)).
136. Greene JA. The materiality of the brand: Form, function, and the pharmaceutical trademark. History and Technology 2013 Jun 1;29(2):210-26.
137. Kesselheim AS, Alpern JD, Stauffer WM. High-cost generic drugs--implications for patients and policymakers. N Engl J Med 2015 Feb 12;372(7):686.
138. Pauwels K, Huys I, Casteels M, Simoens S. Drug shortages in European countries: a trade-off between market attractiveness and cost containment? BMC Health Serv Res 2014;14:438.
139. Kardas-Nelson M. Can (and should) Africa make its own medicines? BMJ 2015;350:h2178.
140. Alpern JD, Stauffer WM, Kesselheim AS. High-cost generic drugs--implications for patients and policymakers. N Engl J Med 2014 Nov 13;371(20):1859-62.
141. Pollack A. Maker of $750-a-Pill Drug Refuses to Lower List Price. The New York Times 2015 Nov 25.
142. Walker J. For Prescription Drug Makers, Price Increases Drive Revenue. The Wall Street Journal 2015 Oct 5.
143. Thomas D. Pfizer Bets $15 Billion on New Sort of Generics. The New York Times 2015 Feb 6.
144. Andrew Ward. GSK chief signals decline of US drug price inflation. Financial Times 2015 Oct 23.
145. Emanuel E. The Solution to Drug Prices. The New York Times 2015 Sep 9.
146. Walker J. Pricing Power Boosts Drug Firms --- Steady increases by companies outpace inflation and often are imposed even when demand falls. Wall Street Journal 2015 Oct 7.
147. Pear R. Senate Aims to Stop Firms From 'Buying Up Drugs and Jacking Up Prices'. The New York Times 2016 Dec 21;US.
148. Onakpoya IJ, Spencer EA, Thompson MJ, Heneghan CJ. Effectiveness, safety and costs of orphan drugs: an evidence-based review. BMJ Open 2015;5(6):e007199.
149. Blankart CR, Stargardt TF, Schreyogg J. Availability of and access to orphan drugs: an international comparison of pharmaceutical treatments for pulmonary arterial hypertension, Fabry disease, hereditary angioedema and chronic myeloid leukaemia. (1179-2027 (Electronic)).
150. Henrard S, Arickx F. Negotiating prices of drugs for rare diseases. Bull World Health Organ 2016 Oct 1;94(10):779-81.

151. Kanters TA, Steenhoek A, Hakkaart L. Orphan drugs expenditure in the Netherlands in the period 2006-2012. Orphanet J Rare Dis 2014;9:154.
152. Kinney J. Health disparities: Exploring the ethics of orphan drugs. Am J Health Syst Pharm 2014 May 1;71(9):692-3.
153. Ridley DB, Regnier SA. The Commercial Market For Priority Review Vouchers. Health Aff (Millwood) 2016 May 1;35(5):776-83.
154. Pollack A. Martin Shkreli's latest plan to sharply raise drug prices prompts outcry. The New York Times 2016 Dec 11.
155. Hyry HI, Cox TM, Roos JC. Saving orphan drug legislations: misconceptions and clarifications. Expert Rev Pharmacoecon Outcomes Res 2016;16(1):111-7.
156. Pollack A, Tavernise S. A Drug Company's Price Tactics Pinch Insurers and Consumers. The New York Times 2015 Oct 5.
157. Bennette CS, Richards C, Sullivan SD, Ramsey SD. Steady Increase In Prices For Oral Anticancer Drugs After Market Launch Suggests A Lack Of Competitive Pressure. Health Aff (Millwood) 2016 May 1;35(5):805-12.
158. FDA. Unaproved Drugs Initiative. FDA gov 2015 December 15 [cited 2016 Jun 2];Available from: URL: FDA.gov
159. Ebbers HC, Pieters TF, Leufkens HG FAU - Schellekens H, Schellekens H. Effective pharmaceuticals regulation needs alignment with doctors. Drug Discov Today 2012 Feb;(1878-5832 (Electronic)).
160. Ahmed I, Kaspar B, Sharma U. Biosimilars: impact of biologic product life cycle and European experience on the regulatory trajectory in the United States. Clin Ther 2012 Feb;34(2):400-19.
161. Mestre-Ferrandiz J, Towse A, Berdud M. Biosimilars: How Can Payers Get Long-Term Savings? Pharmacoeconomics 2016 Jun;34(6):609-16.
162. Vogler S, Zimmermann N, Ferrario A, Wirtz VJ, de JK, Pedersen HB, et al. Pharmaceutical policies in a crisis? Challenges and solutions identified at the PPRI Conference. J Pharm Policy Pract 2016;9:9.
163. Davey M. Australia's poorly applied drug policy wastes $320m a year: study. The Guardian 2015 Jun 21.
164. Meier B. New Dosages of Old Drugs Help Raise Their Prices. The New York Times 2015 Feb 11.
165. Rashidian A, Omidvari AH, Vali Y, Sturm H, Oxman AD. Pharmaceutical policies: effects of financial incentives for prescribers. Cochrane Database Syst Rev 2015;8:CD006731.
166. Hernu R, Cour M, de la Salle S, Robert D, Argaud L. Cost awareness of physicians in intensive care units: a multicentric national study. Intensive Care Med 2015 Aug;41(8):1402-10.
167. Bucx MJ, Landman JJ, van Onzenoort HA, Kox M, Scheffer GJ. A simple intervention to reduce the anesthetic pharmacy budget; the effect of

price list stickers placed on vaporizers. J Clin Anesth 2015 Jun;27(4):307-10.
168. Galizzi MM, Ghislandi S FAU - Miraldo M, Miraldo M. Effects of reference pricing in pharmaceutical markets: a review. Pharmacoeconomics 2011 Jan 1;29(1179-2027 (Electronic)):17-33.
169. Senior M. Heart drug pushes outcome-based pricing plans. NATURE REVIEWS DRUG DISCOVERY 2015 Oct 1;14(1474-1784 (Electronic)):665-7.
170. Houy N, Jelovac I. Drug Launch Timing and International Reference Pricing. Health Econ 2015 Aug;24(8):978-89.
171. Kay M. India moves closer to regulating prices of patented drugs. BMJ 2014;348:g1244.
172. Vogler S, Vitry A, Babar ZU. Cancer drugs in 16 European countries, Australia, and New Zealand: a cross-country price comparison study. Lancet Oncol 2016 Jan;17(1):39-47.
173. Mueller MT, Frenzel A. Competitive pricing within pharmaceutical classes: evidence on "follow-on" drugs in Germany 1993-2008. Eur J Health Econ 2015 Jan;16(1):73-82.
174. Fuller RL, Goldfield N. Paying for On-Patent Pharmaceuticals: Limit Prices and the Emerging Role of a Pay for Outcomes Approach. J Ambul Care Manage 2016 Apr;39(2):143-9.
175. Andrew Ward. Novartis chief backs shake-up in drug pricing. Financial Times 2015 Jun 7.
176. Jonathan Ford. Drug pricing comes under the microscope. Financial Times 2015 Aug 16.
177. Collins M, Latimer N. NICE's end of life decision making scheme: impact on population health. BMJ 2013 Mar 21;346(1756-1833 (Electronic)).
178. Boseley S. NHS to rein in cost of Cancer Drugs Fund. The Guardian 2015 Jan 8.
179. Boseley S. Cancer drug companies cut prices to win NHS approval. The Guardian 2016 Aug 18.
180. Loftus P. New Push Ties Cost of Drugs to How Well They Work. The Wall Street Journal 2015 May 26.
181. Boseley S. Patients suffer when NHS buys expensive new drugs, says report. The Guardian 2015 Feb 19.
182. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. HEALTH TECHNOLOGY ASSESSMENT 2015 Feb;19(2046-4924 (Electronic)).
183. Andrew Ward. Expensive drugs cost lives, claims report. Financial Times 2015 Feb 19.
184. Bach P. Why Drugs Cost So Much. The New York Times 2015 Jan 15.

185. Ferrario A, Kanavos P. Dealing with uncertainty and high prices of new medicines: a comparative analysis of the use of managed entry agreements in Belgium, England, the Netherlands and Sweden. Soc Sci Med 2015 Jan;124:39-47.
186. Andrew Ward, Ralph Atkins. Novartis chief looks to blockbuster drugs to counter profits drop. Financial Times 2015 Jul 21.
187. Loftus P, Anna WM. Insurers Push to Tie Drug Prices to Outcomes. Wall Street Journal 2016 May 11.
188. Loftus P. New Novartis Heart-Failure Drug Should Cost 17% Less, Research Group Says. The Wall Street Journal 2015 Sep 11.
189. Inagaki K. Japan to step up drug price reviews in effort to cut health spending. FT.com 2016 Dec 20.
190. Beall RF, Nickerson JW, Attaran A. Pan-Canadian overpricing of medicines: a 6-country study of cost control for generic medicines. Open Med 2014;8(4):e130-e135.
191. Rockoff JD. Drug-Price Rises Stick Stubbornly. Wall Street Journal 2016 Nov 28.
192. Rockoff JD. Business News: Drug Firms' Pricing Power Sticks. Wall Street Journal 2016 Nov 28.
193. Ward A. World's first malaria vaccine gets green light from Europe. Financial Times 2015 Jul 24.
194. Kollewe J. First malaria vaccine given green light by European regulators. The Guardian 2015 Jul 24.
195. The Editorial Board. Hope for a Malaria Vaccine. The New York Times 2013 Oct 13.
196. Husereau D, Dempster W, Blanchard A, Chambers J. Evolution of drug reimbursement in Canada: the Pan-Canadian Pharmaceutical Alliance for new drugs. Value Health 2014 Dec;17(8):888-94.
197. Alpern JD, Song J, Stauffer WM. Essential Medicines in the United States--Why Access Is Diminishing. N Engl J Med 2016 May 19;374(20):1904-7.
198. Owoeye OA. Compulsory patent licensing and local drug manufacturing capacity in Africa. Bull World Health Organ 2014 Mar 1;92(3):214-9.
199. Chatterjee C, Kubo K, Pingali V. The consumer welfare implications of governmental policies and firm strategy in markets for medicines. J Health Econ 2015 Oct 22;44:255-73.
200. Andrew Jack. Free Lunch: No free drugs. Financial Times 2015 Oct 1.
201. The Editorial Board. No Justification for High Drug Prices. The New York Times 2015 Dec 20.
202. Pollack A. Skin Drugs Lead Way in Price Rise, Study Says. The New York Times 2015 Nov 26.
203. Durham D. Inflated Drug Prices. The New York Times 2015 Sep 14.

204. Andrew Ward. GlaxoSmithKline chief questions rivals' chasing M&A deals. Financial Times 2015 May 11.
205. Cvek B. Nonprofit drugs as the salvation of the world's healthcare systems: the case of Antabuse (disulfiram). Drug Discov Today 2012 May;17(9-10):409-12.
206. Ft view. Test new financial models for antibiotics. FT.com 2016 Sep 18.
207. Ward A. GSK to relax intellectual property and help developing world. Financial Times 2016 Mar 31.
208. World Bank. New Country Classifications. data worldbank org 2015 February 7Available from: URL: data.worldbank.org
209. Kollewe J. GlaxoSmithKline to lower drug prices in poorer countries. The Guardian 2016 Mar 31.
210. McKellar MR, Fendrick AM. Innovation of novel antibiotics: an economic perspective. Clin Infect Dis 2014 Oct 15;59 Suppl 3:S104-S107.
211. Rex JH, Outterson K. Antibiotic reimbursement in a model delinked from sales: a benchmark-based worldwide approach. Lancet Infect Dis 2016 Apr;16(4):500-5.
212. Andrade LF, Sermet C, Pichetti S. Entry time effects and follow-on drug competition. Eur J Health Econ 2016 Jan;17(1):45-60.
213. Crow D. Merck launches hepatitis C drug price war. Financial Times 2016 Jan 29.
214. Niezen MG, Stolk EA, Steenhoek A, Uyl-De Groot CA. Inequalities in oncology care: Economic consequences of high cost drugs. Eur J Cancer 2006 Nov;42(17):2887-92.
215. Howard DH, Chernew ME, Abdelgawad T, Smith GL, Sollano J, Grabowski DC. New Anticancer Drugs Associated With Large Increases In Costs And Life Expectancy. Health Aff (Millwood) 2016 Sep 1;35(9):1581-7.
216. Aggarwal A, Sullivan R. Affordability of cancer care in the United Kingdom â€“ Is it time to introduce user charges? Journal of Cancer Policy 2014;2(2):31-9.
217. Dolgin E. Big pharma moves from 'blockbusters' to 'niche busters'. Nat Med 2010 Aug;16(8):837.
218. Ubel PA, Bach PB. Copay Assistance for Expensive Drugs: A Helping Hand That Raises Costs. Ann Intern Med 2016 Dec 20;165(12):878-9.
219. Pear R. Obama Administration Seeks Ways to Rein In Pharmaceutical Costs. The New York Times 2015 Nov 21.
220. Andrew Ward. UK watchdog accuses Pfizer of inflating anti-epilepsy drug price. Financial Times 2015 Aug 6.
221. Pear R. Obama Administration Plans Forum on High Drug Prices. The New York Times 2015 Nov 4.

4.2.8.1 Supporting Information

S1 PRISMA 2009 checklist - Jun2017.docx

4.3 Evaluation

4.3.1 Strengths

The strength of this paper is the comprehensive overview of the global drugs market, which is required due to the international character of the pharmaceutical industry and the possible applicability of local solutions. The mixed-sources approach, of both scientific articles and newspaper articles, allowed for a comprehensive view from different angles and markets.

4.3.2 Limitations

One of the weaknesses of this paper is the selection methodology. The difficulty with newspapers is that they are not as easily searched as scientific literature. Different search engines were used with customized search terms, to make sure a reasonable number of articles was retrieved for screening. This may have given a skewed representation of the field, as one newspaper may have been better indexed and more searchable than another, making one range of opinions more prominent than others. Given that the final number of included articles (19, 20, 14 and 15 for the FT, NYT, Guardian and WSJ, respectively) appear reasonably balanced, this effect is not likely to be significant.

Another shortcoming of the search methods is the newspaper selection. By selecting these four newspapers, we mostly obtained a Western perspective on the drug market. To gain insights from the Asian, African and South American markets, we could have selected different newspapers that would yield better information on the challenges and solutions in those settings. This would have broadened the scope further, though, complicating the overview and the recommendations.

Indeed, the main weakness of this study is the large scope. It is impossible to create a complete overview of all the relevant contributors of the market dynamics in many different countries without losing a clear problem definition and possible solutions. On the one hand, we would like to research and present a comprehensive overview of the worldwide drug market, with lessons that can be applied regardless of local landscapes and to find solutions that are universally applicable. Considering the international character of drug development and production, this would fit the subject well. That would however mean that we would make an overview of all dynamics in all countries, which would be so fine-grained that it is easy to get lost in the details.

Assessing the quality of each paper was not always possible. The method of assessing would then have to be different for newspaper articles, model-based approaches for policy impacts, historic analyses and case-control studies on policy effects. Though these publications are not related to a pre-registered database, and the results are not quantifiable, we could have assessed the quality of studies that were included.

4.3.3 Appropriateness of the methodology

Complementing scientific source material with newspaper articles gave us insight in the real-world implications of abstract concepts. The newspaper stories tended to focus on the patient perspective, whereas scientific papers often looked at market dynamics and policy. So including popular media outlets was appropriate and enriched the SR.

4.3.4 Lessons learned

If we were to re-write this review, we might consider including news sources from other countries, to reflect the growing importance of developing countries to the drug market. We would also search for a universal search engine that indexes all journals, so that consistent search criteria and indexing can be used. Ideally, we would have two reviewers select articles and discuss the differences.

This review creates an overview of the landscape of drug pricing, which could be the basis for future research. A new SR could focus on one individual policy intervention, and summarize all available data on that specific intervention. Having a broader understanding of the interwoven drug market and policies will help with assessing what unintended consequences could be expected.

Before meta-analyses on policy are possible, let alone meta-regressions and network analyses, significantly more data needs to be generated. With the currently available data, a meaningful integration of the effect of policy is not possible, as the details of interventions are too complex and different between countries. Furthermore, the settings in which they work are not often comparable. For now, policy interventions will have to be summarized by scoping reviews and systematic reviews without data integration.

4.4 Letters

After the publication of this review, we found two papers that were published after ours with conclusions we did not find valid. It is essential in evidence generation that data is discussed by methodological experts and content matter experts, to make sure the quality of each paper is assessed properly. In these two cases, we sent a letter to the journal that published the studies to debate the findings and allow the original authors to respond.

This was possible because we had written an SR, which showed that we had taken all relevant literature into account to form a reliable opinion. This made us the content matter experts that were needed for the dissection of findings. Our contribution to the field gave us the intellectual credibility to debate the findings of others in scientific journals.

4.4.1 Assessing Pharmaceutical Research and Development Costs

4.4.1.1 Context

This letter was submitted to JAMA internal medicine, in response to a primary research article. In that paper, the authors discussed the costs of bringing a single drug in oncology to the market. They used R&D spending reported in filings to the US Securities and Exchange Commission filings by companies with one single agent in research, and concluded that their findings could be extrapolated to all companies, even to research in other indications. We wrote a letter debating the methodology and validity of the conclusions.

Prasad, V., & Mailankody, S. (2017). Research and Development Spending to Bring a Single Cancer Drug to Market and Revenues After Approval. *JAMA Internal Medicine*, 177(11), 1569. https://doi.org/10.1001/jamainternmed.2017.3601

4.4.1.2 Full text

A preliminary version of this chapter was published in JAMA internal medicine.

van der Gronde, T., & Pieters, T. (2018). Assessing Pharmaceutical Research and Development Costs. *JAMA Internal Medicine*, 178(4), 587. https://doi.org/10.1001/jamainternmed.2017.8706

Assessing Pharmaceutical Research and Development Costs

To the Editor In their Original Investigation published in a recent issue of JAMA Internal Medicine, Prasad and Mailankody[1] examine the costs of bringing a single drug with an oncological indication to the market. They conclude that 10 selected companies had a median investment cost of $648 million for the development of their drug, whereas the median revenue in 4 years on average was $1658.4 million. This further reinforces the notion that there are large profits to be made with drug development, and that current pharmaceutical drug prices are unrelated to the actual costs for research and development.

There are, however, some caveats to the analysis by Prasad and Mailankody.[1] The authors only selected companies that had no other drugs on the market at the time of filing for US Food and Drug Administration approval, meaning only small companies. It is possible that small companies have a leaner organization and can work more efficiently than larger pharmaceutical companies can, so the generalized conclusion that it costs $648 million to bring a drug to market is, even if regarding only the market for oncology drugs, a potentially inaccurate extrapolation of their findings. The costs of bringing a drug to the market might be lower for a small company than for a larger company.

Additionally, this analysis only considered companies that had been successful in pursuing market authorization for their product. Next to those, there were likely competitors with similar projects that did not yield a successful product. The costs of those unsuccessful investments should also be taken into account to make a fair estimation of the costs of development.

The study outcomes, however, are valuable and contribute to a larger movement we recently described in an article on drug pricing.[2] Prices of drugs are not related to their development costs but to what the market will pay for them. The double-digit profit marginscommonlyseen in the industry are a clear result of this. Without adequate regulation to control prices in the United States, drug prices can be expected to rise further.We agreewithPrasadandMailankody1 thatmore transparency is needed to clarify the discrepancy between investments and prices in the pharmaceutical industry.

This analysis[1] highlights that society is overpaying for the products developed by drug companies. With more regulation on pricing, prices can be brought down significantly without discouraging innovation.

Toon van der Gronde, PharmD

Toine Pieters, PhD

Author Affiliations: Department of Pharmaceutical Sciences, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Utrecht, the Netherlands.

Corresponding Author: Toine Pieters, PhD, Utrecht University, Pharmaceutical Sciences, Princetonplein 5, Utrecht 3508AD, the Netherlands (t.pieters@uu.nl).

Conflict of Interest Disclosures: None reported.

1. Prasad V, Mailankody S. Research and development spending to bring a single cancer drug to market and revenues after approval. JAMA Intern Med. 2017;177 (11):1569-1575. doi:10.1001/jamainternmed.2017.3601

2. Gronde TV, Uyl-de Groot CA, Pieters T. Addressing the challenge of high-priced prescription drugs in the era of precision medicine: a systematic review of drug life cycles, therapeutic drug markets and regulatory frameworks. PLoS One. 2017;12(8):e0182613. doi:10.1371/journal.pone.0182613

4.4.2 Response to proposal for a novel cancer drug pricing model

4.4.2.1 Context

We submitted this letter in response to a paper authored by one of our previous co-authors, in which the authors proposed a new drug pricing mechanism. The new model used R&D costs for a single drug, the size of the target population, the remaining time that the drug is still on patent, and a profit margin related to the anticipated level of benefit. We debated the impact this new pricing model would have on the highly complex pharmaceutical market, what this would incentivise and how this could be improved.

Uyl-de Groot, C. A., & Löwenberg, B. (2018). Sustainability and affordability of cancer drugs: a novel pricing model. *Nature Reviews Clinical Oncology*, (May). https://doi.org/10.1038/s41571-018-0027-x

4.4.2.2 Full text

A preliminary version of this chapter was published in Nature reviews clinical oncology.

van der Gronde, T., Leufkens, H. G., & Pieters, T. (2018). Response to proposal for a novel cancer drug pricing model. *Nature Reviews Clinical Oncology*. https://doi.org/10.1038/s41571-018-0062-7

Response to proposal for a novel cancer drug pricing model

Toon van der Gronde, Hubertus G. Leufkens and Toine Pieters

In their recent News & Views article (Sustainability and affordability of cancer drugs: a novel pricing model. Nat. Rev. Clin. Oncol. 15, 405–406 (2018))[1], Uyl- De Groot and Löwenberg outline a new universal algorithm for setting the price of new drugs in oncology. Their ambitious proposal is intended to standardize a complicated and fragmented pricing process. The international drug market is dynamic and diverse[2]. Having a universal pricing mechanism could indeed be helpful in addressing the imbalances in drug pricing and improve access to medicines for many patients, including those with non-oncological diseases. However, we would like to point out some thoughts on the proposed algorithm.

First, the scope of the algorithm seems to be limited to the USA and European Union (EU). Outrage about high drug prices and the associated limitations in access to potentially life- saving drugs are certainly not limited to these regions. Importantly, most of the economic and population growth in the next few decades, and subsequently the burden of disease, will be in developing areas of Asia and Africa. Therefore, to have a truly universal framework, these regions would have to be included. Furthermore, the correction factors for countries with varying levels of economic development should be based on a more sophisticated tool than the one proposed, based on gross domestic product per capita equivalent costs per disability- adjusted life year averted[3].

Second, all parameters of the algorithm carry inherent risks of unwanted stakeholder behaviour, by both industry and payers. Acknowledging research and development (R&D) costs seems reasonable, but this does not necessarily encourage improvements in clinical benefit. Even if the price of R&D could be calculated objectively (the analysis cited by the authors[4] is already highly debated[5–7]) and if governments succeed in transnationally coordinating the reimbursement of R&D costs, this could stimulate spending on R&D with very little incentive to bring meaningful improvements in the quality of life of patients.

Third, years left on patent is a parameter that provides ample opportunity for interest-based manoeuvring. Patents are frequently extended or renewed based on additional research efforts (such as extension of a patent after the completion of research in paediatric populations[8,9]) or changes in formulation, thus making estimates of remaining patent life unreliable. Furthermore, the 3-year evaluation period would incentivize producers to delay the launch of any new drug indications until after the recalculation, so that the target population is smaller and, therefore, the accepted drug price will be higher. Similarly, such measures would encourage companies to introduce new indications as quickly as possible following recalculation, in order to maximize financial benefit at the expense of both patients and health-care providers.

In conclusion, the development of a universal framework to guide something as complex as the drug pricing process is an admirable undertaking that merits ample debate. Any model that is adopted will drive stakeholder behaviour towards optimally satisfying their own interests. The key challenge is, therefore, to align the interests of stakeholders with those of the general public. Even the best available model, value- based pricing, can lead to unacceptable outcomes and limitations in access. This scenario is seen in the recent debate around hepatitis C products, which have superb clinical benefits but are unaffordable for the population2. The proposed algorithm is a laudable initiative, but requires careful assessment to ensure the provision of a framework that helps streamline the pricing process and also ensures equal access to new drugs for patients around the world.

There is a reply to this letter by Uyl- de Groot, C. A. & Löwenberg, B. Nat. Rev. Clin. Oncol. https://doi.org/10.1038/s41571-018- 0063-6 (2018).

Toon van der Gronde[1,2], Hubertus G. Leufkens[2] and Toine Pieters[2]*

[1]AstraZeneca, Cambridge, United Kingdom.
[2]Utrecht Institute for Pharmaceutical Sciences (UIPS) Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, Netherlands.
*e- mail: t.pieters@uu.nl https://doi.org/10.1038/s41571-018-0062-7

1. Uyl- de Groot, C. A. & Löwenberg, B. Sustainability and affordability of cancer drugs: a novel pricing model. Nat. Rev. Clin. Oncol. 15, 405–406 (2018).
2. Van der Gronde, T., Uyl- de Groot, C. A. & Pieters, T. Addressing the challenge of high- priced prescription drugs in the era of precision medicine: a systematic review of drug life cycles, therapeutic drug markets and regulatory frameworks. PLOS ONE 12, e0182613 (2017).
3. Marseille, E. et al. Thresholds for the cost- effectiveness of interventions: alternative approaches. Bull. World Health Organ. 93, 118–124 (2015).
4. Prasad, V. & Mailankody, S. Research and development spending to bring a single cancer drug to market and revenues after approval. JAMA Intern. Med. 177, 1569 (2017).
5. Ledley, F. D. Methods used to assess pharmaceutical research and development costs. JAMA Intern. Med. 178, 589 (2018).
6. DiMasi, J. A. Assessing pharmaceutical research and development costs. JAMA Intern. Med. 178, 587 (2018).
7. van der Gronde, T. & Pieters, T. Assessing pharmaceutical research and development costs. JAMA Intern. Med. 178, 587–588 (2018).
8. Nelson, R. E. et al. Patent extension policy for paediatric indications: an evaluation of the impact within three drug classes in a state Medicaid programme. Appl. Health Econ. Health Policy 9, 171–181 (2011).
9. The European Medicines Agency. Rewards and incentives for paediatric medicines. EMA http:// www.ema.europa.eu/ema/index.jsp?curl=pages/ regulation/general/general_content_000607. jsp&mid=WC0b01ac0580925b1c (2018).

5 Gene doping: an overview and current implications for athletes

This chapter looks at a third solution to find the relevant literature. The question was specifically to re-assess something that had been established five years prior, so that was our starting point. We used the selection of key proteins from that earlier publication, focussing on what had happened in the five years since. This meant that we used eleven separate search strings to find the right literature.

This chapter also looks at how to synthesise data and comes up with a scoring method that allows for a conclusion. As discussed in paragraph 2.2.6, presenting results in such a way that they can be compared is where SRs add value to the field. As there is no easy and established way to calculate this for non-comparable studies, we pioneered a method to estimate this.

5.1 Context

This SR consists of eleven related searches, with one overarching goal. Five years prior to this SR, authors of another review had suggested that gene doping would enter the sports arena within five years. This review examined that question: what is the probability that gene doping had been used since then?

To answer our question, we looked at the state of the science of the proteins that are most likely to be abused for gene doping. The earlier review mentioned several likely targets, and other reviews mentioned more options. This allowed us to use these as search terms. Eleven targets were chosen: erythropoietin (EPO), insulin-like growth factor (IGF), growth hormone (GH), myostatin, vascular endothelial growth factor (VEGF), fibroblast growth factor (FGF), endorphin and enkephalin, α-actinin 3, peroxisome proliferator-activated receptor-δ (PPARδ) and cytosolic phosphoenolpyruvate carboxykinase (PEPCK-C).

This review focussed on those eleven proteins to determine the most likely targets and the likelihood of gene doping use. As gene doping is the abuse of gene therapy, that term was also included in the search. We used a simple search string, which was adjusted for each protein: “gene doping” OR “gene therapy” AND [target]. This string was used to search PubMed, Embase and Scopus to provide a reliable overview. Literature was limited to the last five years, to reflect the research

question. In vitro research was excluded, to focus solely on animal and in-human research.

To allow for an overall estimation of the likelihood of gene doping abuse, we scored each protein for an attribute that would make it more tempting to use. We used a scale ranging from ++ (very high) to - - (very low) for four categories: potential benefits, experience in gene therapy, risk control and chance of undetected use. This allowed us to calculate a composite score of likelihood of abuse. This was based on the found literature, by agreement between the authors.

In addition, we discussed the basic mechanisms of gene therapy, potential vectors to use for gene therapy, the current understanding of the risks of gene doping, direct and indirect detection methods. This all helped contextualize the research field for a better understanding. The PRISMA reporting methodology was used to ensure quality of reporting.

5.2 Full text

A preliminary version of this chapter was published in the British Journal of Sports Medicine.

van der Gronde, T., de Hon, O., Haisma, H. J., & Pieters, T. (2013). Gene doping: an overview and current implications for athletes. *British Journal of Sports Medicine*, 47(11), 670–678. https://doi.org/10.1136/bjsports-2012-091288

Gene Doping: An Overview and Current Implications for Athletes

Toon van der Gronde[1], Olivier de Hon[2], Hidde J. Haisma[3], Toine Pieters[1, 4]

[1]*Department of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Utrecht, the Netherlands.*

[2]*Anti-Doping Authority the Netherlands, Capelle aan den IJssel, the Netherlands*

[3]*Department of Pharmaceutical Gene Modulation, Groningen Research Institute of Pharmacy, Groningen University, the Netherlands*

[4]*Community genetics (EMGO), VU Medical Centre, Amsterdam, the Netherlands*

Correspondence to:

Prof. Dr. Toine Pieters

E-mail address: t.pieters@uu.nl

Keywords: Athletics, doping, evidence-based reviews, genetics, elite performance

5.2.1 Preface

Gene therapy is a promising new treatment modality but can be misused to improve athletic performances. To date, no detection method has been approved for gene doping and this is a problem for the anti-doping sports community. The primary aim of this review is to present the current status of gene doping and detection methods in order to answer the following pressing question: Is it likely gene doping will be used at major sporting competitions, and if so, what are the odds of effective detection?

5.2.2 Introduction

Gene therapist Ted Friedmann and multiple Olympic gold medallist Johann-Olav Koss were the first to describe the possibility of misusing the techniques and experiences of gene therapy in the athletic arena. (1) In 2006, before the Turin Winter Olympic games, the president of the World Anti-Doping Agency (WADA), Dick Pound, called gene doping 'the new threat that is now a reality'. (2) Although Pound did not expect gene doping to pose a problem in Turin, he indicated it could be a problem at the Summer Games, two years hence in Beijing. In fact, the problem did

not materialize in China, in 2008, nor at the London 2012 Olympics, as far as the then available detection measures could determine. Yet again, we have to operate on the assumption that there may be athletes out there willing to test gene doping at the 2016 Rio de Janeiro Olympics. After all, an Olympic gold medal means considerable social and economic benefit. (3-7) Historical doping control statistics show that somewhere between 1.1 and 2% of all athletes test positively for doping. (8) The real number of doping users is expected to be higher, despite the fact that the governing bodies of sport place immense pressure on athletes by a strict liability rule that makes them responsible for everything in their bodies. (9) Although the detection of doping is constantly improving, it generally trails actual practice. (4,9) In a 2006 review, Haisma and De Hon stated that gene doping was likely to enter sports within five years. (10) Given that gene therapy techniques have improved considerably, the likelihood of gene doping has increased ever since. (9-12)

Today, most gene therapy studies examine hereditary diseases and cancer. (13,14) Gendicine (Recombinant Human Ad-p53 Injection) and Glybera (alipogene tiparvovec) are the first approved gene therapy products for human use in the USA and the EU respectively. Gendicine is designed to place a p53 gene in cancer cells to inhibit cell growth and the Glybera gene therapy has been approved for treatment of life threatening pancreatitis attacks in patients with lipoprotein lipase deficiency (LPL). (15-17)

The proteins selected for this systematic review include those reviewed by Haisma and De Hon in 2006. (10) Additional proteins were included if determined to be likely candidates for misuse in (potential) gene doping because of their physiological effects and current status in anti-doping regulations, or the possibility of gene isolation and manipulation using techniques available in gene therapy. The final list of selected proteins are erythropoietin, insulin-like growth factor, growth hormone, myostatin, vascular endothelial growth factor, fibroblast growth factor, endorphin and enkephalin, α actinin 3, peroxisome proliferator-activated receptor-delta and cytosolic phosphoenolpyruvate carboxykinase. These proteins are thus the most relevant for the following systematic review, but they are not an exhaustive list of all possible proteins with an impact on athletic performance.

First, gene therapy and the risks and the safety related to gene doping is discussed. Subsequently, the properties and targets of the aforementioned therapeutic proteins are reviewed, as well as their current (pre-)clinical status. Next, animal models, gene therapy, and gene doping are considered. Each protein was scored for potential benefits to athletes, experience in gene therapy, controllability of the risks, and the chance of using the protein without detection. The scoring results were used to consider the degree to which the technique and protein was likely to be misused in sports, now or in the near future. Finally, detection methods including direct and indirect, as well as animal use of gene doping are addressed.

5.2.3 Methods

A general literature search was conducted to identify the most relevant proteins based on their current gene doping potential using articles from Pubmed, Scopus and Embase published between 2006 and 2011. The final list of 11 proteins included those reviewed by Haisma and De Don in 2006. A systematic review of the 11 proteins was then conducted (10) by using the search terms "gene doping'' OR ''gene therapy'' AND [protein]. Exclusion criteria were:

- Research before 2006
- In-vitro research
- articles published in languages other than English

Articles were judged by their title for relevance, i.e. whether they concerned the appropriate protein and in adequate detail. If the title did not provide adequate certainty for inclusion, the abstract (if available) was judged. Articles describing the use of the protein in gene therapy or gene doping and review articles were preferred; although, primary research articles were also included in the sample. For each protein, articles were selected until a saturation point had been reached, i.e., additionally identified articles offered no new information (see figure 1).

The search for eligible articles for the systematic review was completed on December 31, 2011 and included articles published from January 2006 through December 2011. Articles published before January 2006 were deemed to be included in a previous review and only those providing context that was critical to the analysis were added. (10) For each of the 11 identified proteins, an average of ten articles were

selected. As a supplement to the systematic search, a small number of important publications published before 2006 were added.

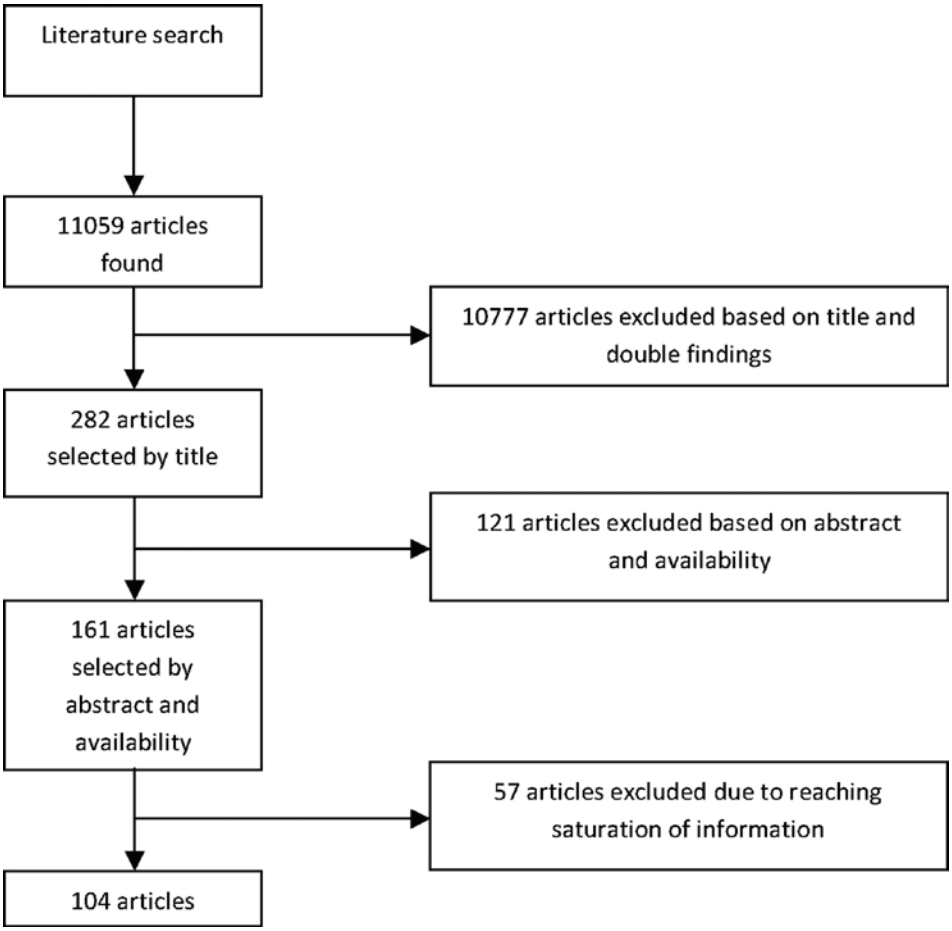

Figure 1. This flow chart shows the overall selection process of the articles. More information can be found in the appendix.

5.2.4 From gene therapy to gene doping

One of the most difficult steps in gene therapy is delivering the gene into host cells. (12) Three major techniques used for delivering genes are injecting naked DNA, viruses or modified cells. (3)

Direct injection of DNA into the target tissue. Initially, a desired gene can be produced in bacterial plasmids and then purified. Next, the gene can be directly injected into the target tissue. Unfortunately, direct injection of DNA is not very effective due to limited uptake and nuclear

translocation (although electroporation of the target tissue increases the uptake(18)). However, it is safer than using viral vectors, since there are generally fewer immune responses(11,13,19), and considerably cheaper than other gene-transfer options. (20)

Introducing genetic material using a virus. Genetic material can be delivered to a target tissue by using a viral vector. Viruses have evolved to efficiently transfect cells with their genetic information and multiply, which makes them ideal for use in gene therapy. To prevent the virus from replicating, the viruses are tamed (all DNA or RNA coding for proteins allowing the virus to multiply and escape infected cells is removed) and therapeutic genes are inserted. Because of this inhibited replication, the viruses are less immunogenic. (9,12,14,19) The modified virus may be injected intravascularly or directly into a target tissue, or inhaled. Injection into the target tissue limits gene expression to the injection site, whereas intravascular injection usually results in systemic expression. (10,21) Inhalation is used if the lungs are the target tissue. (10,17,22) The viral vector contains a promoter that allows the inserted gene to be transcribed and translated, thus yielding the desired protein. (10)

Ex-vivo gene therapy. For ex-vivo gene therapy, stem cells are removed as in the case of patients with SCID and a therapeutic gene is introduced in vitro. (9,23,24) The genetically modified stem cells are then injected back into the patient's bone marrow. This can be done using plain DNA (with or without liposomes), a viral vector with electroporation, or with a gene gun. (9,12) The technique allows for limited screening and sorting of the cells before re-injection, which increases efficacy and safety. Disadvantages are low efficiency and increased cost. (13,24)

Depending on the target tissue, the gene and the desired duration of transgene expression, multiple vectors can be used. (17,25,26) The most important properties are displayed in table 1.

Table 1. This table presents an overview of the properties of the most used viral vectors in gene therapy. The column "Relative use" describes the use in gene therapy trials. Also see the Gene Therapy Clinical Trials Worldwide database. ds: double strand

Virus	**Relative use (%)**	**Package capacity (kb)**	**Tissues**	**Advantages**	**Disadvantages**
Adenovirus	24(12,27)	30 (12,27, 28) of ds DNA (17,26)	Wide range(19, 29)	- No cell division required (17,29) - Experience (28) - No integration (9,12,17,19,28,30) - Large capacity(14,26)	- Short expression (12,29-31) - Immunogenic when delivered systemically (14,19,26,27,32)
Classic Retrovirus	20	8(19) of ss RNA(17)	Wide range(33)	- Experience	- Lack of specificity(33) - Once caused leukemia (14,33)
Vaccinia virus	6	25 of ds DNA(34)	Almost all cells (34)	- Much experience - Antiviral agent exists - Well storable - No integration(34)	- High immune response - short effect(34)
Adeno-associated virus (AAV)	5	4,5 (12,14,19,22,28) of ss DNA	Neurons, muscle, brain, liver and	- Not pathogenic - Long effect (14,17,19,28,29, 35,37,38)	- Requires a helper virus to integrate (14,22,31)

		(17,26,30)	hematopoietic cells (19,28,35,36)	- No cell division required (14) - Almost no (<1%) integration(14)	- Lack of experience (14)
Herpes Simplex virus (HSV)	3	Up to 150 (12,28, 39) of ds DNA(17)	Primary afferent neurons, epithelial and mucosal cells (29,35, 39,40)	- Can be present in a latent state (12,39,41) - Low immune response(28) - No integration (17,40)	- Can be present in a latent state (12,39,41) - Neurotoxicity and cytohepathic problems (31)
Lentivirus	2	8 (12,19, 26) of ss RNA (17,26)	Macrophages, neurons (12,19, 42)	- Experience (35) - No cell division required (14) - Low immune response (25)	- Random integration (12,19,33) - Errors in transcription (31)
Plasmid vectors	19	10 (12,43)	Wide range	- Less immunogenic - Less oncogenic -No recombination with other viruses (12,31,35,43) - Cheaper (44) - No need to work with viruses(12)	- Varying expression (12,31) - Low efficacy (14,19,44)

Box 1

Risks of gene doping

Like all new medicines, gene therapy presents unsolved problems.(36) Until they are solved, large-scale use of gene therapy in the clinic is ruled out. Contrary to gene therapy and by implication of its illegal character, gene doping is not bound to safety regulations.(13)

This box presents an overview of all risks involved with the application of gene doping in sports. Given the risks at this point, gene doping should be prohibited on safety grounds alone.(4,27,40)

Gene silencing

One limit to the effectiveness of gene therapy is gene silencing; thus, even when the target tissue is infected, it might not express the inserted gene. (22)

Immune reaction

Both the virus used and the protein itself can cause an immune reaction. How to handle this reaction appropriately is not completely clear.(9,13,14,32,76,86,98) The immune reaction against the protein can also induce a response against the endogenous protein, as happened with EPO in macaques resulting in anaemia.(9,10,19,36,54)

Integration

Though not all viruses integrate, those that do can present problems. Splitting up a tumour suppressor gene, or worse, increasing production of a proto-oncogene, can lead to cancer.(3,6,11,13,22,27,55,137) It is estimated that about one in every 10.000 retroviral insertions might be dangerous, (34) and one in every 10^9 could induce cancer.(20)

Infection of germ cells

The danger of the infection of germ cells with gene therapy also exists. This would transfer exogenous genes to future generations.(45) Though it is explicitly prohibited to target cells that reproduce,(10) and it is not likely gene therapy (which is not aimed at germ cells) would cause an infection of germ cells, this risk should be strictly monitored.(45)

Expression

Expression of gene therapy is hard to control, and overexpression could be dangerous. In addition to the effects of the protein itself, toxicity by accumulation is also dangerous.(6) Also, if a cell producing the desired product is infected by another virus, this could lead to overexpression of the protein.(12,30) Expression is controllable by an inducer drug (for example, doxycycline , which is approved for human use), but since this is detectable, it is less likely to be used for gene doping.(18,47)

Storage and usage

Gene therapy would require good storage and it is questionable whether those wishing to abuse it are knowledgeable about proper handling procedures.(41) Since professionals face difficulties with the non-linear dosage-expression relationship, it is likely those who haven't been thoroughly educated could do worse.(13,41) Even more dangerous would be an attempt to produce gene doping in rogue laboratories, leading to unsafe products.(10,36)

Long term

Gene therapy is a new technology. Only short-term studies have been conducted, which means that the long-term effects are not yet clear.(85) There may still be problems with gene therapy products that simply haven't been identified yet.(3,19,27,36,45,135)

It should be emphasised that the above-listed risks are only the foreseeable risks, as all the information was obtained in regulated and controlled settings.(13) The unknown risks present a much larger problem, because they are far more difficult to anticipate.(3)

Of the three gene-therapy delivery methods described above, in-vivo viral gene transfer is the most successful method for now. (12) In general, the benefits are efficacy and low cost; downsides are immune responses and poor controllability of integration and expression. (19)

Ex-vivo viral gene transfer can be used to insert a gene to produce a desired protein, or increase or inhibit the transcription of an already present gene by influencing promoters. On the other hand, gene expression can also be prevented using antisense RNA sequences. RNA

sequences bind to the original gene, prevent translation and cause destruction by the RNAse H or the siRNA pathway. Even splicing patterns can be altered by blocking splicing recognition sequences, thus allowing for the inclusion or exclusion of specific exons. (14,19,45)

5.2.4.1 Uses of gene therapy.

Gene therapy can be used to treat a variety of illnesses. It may be applied to weaken or kill cancer cells by triggering apoptosis, to enable target cells to produce a protein that otherwise has to be administered or to upregulate the production of a specific protein. (3,31,34) Though a couple of gene therapy products have been marketed outside China to date, at least 1,843 gene therapy trials have been conducted worldwide with thousands of patients suffering from cancer diseases, cardiovascular and neurological diseases and a range of other diseases. (10,15,19,34) Early clinical trials in Europe and the US had limited results and even fatalities were reported from gene therapy; (9,46,47) however, examples of successful gene therapy include treatment of SCID-X1 and Leber's congenital amaurosis. (17) The proof of concept of various transfer strategies in gene therapy shows that we are at least at the beginning of a gene therapy revolution for patients with monogenic diseases. (16,17,48)

5.2.4.2 Regulations.

Regulatory oversight is stricter for gene therapy trials than for most clinical trials due to the potential risks. (35) Since the first documentation of fatalities during gene therapy, regulations have been tightened, primarily in Europe and the US. (46,49) In the US and Europe gene therapy is only allowed in cells that do not reproduce, preventing gene therapy from affecting following generations. (10)

5.2.4.3 Gene doping.

When gene therapy is used to increase the performance of a healthy person, it is considered gene doping by WADA. (4,12) Gene doping presents the same advantages over regular doping as gene therapy does for regular medicine, but detection of gene doping is more difficult. (4,12) Since gene doping is a powerful tool to boost performances, it may have a significant impact on the professional sports world. (4,13)

Gene doping has been prohibited by the International Olympic Committee (IOC) since 2003. In 2004, WADA took responsibility for publishing the Olympic doping list, and they added gene doping. (4,10)

The following new methods with the potential to enhance sport performance, are prohibited:

1. The transfer of nucleic acids or nucleic acid sequences;

2. The use of normal or genetically modified cells. (50)

Together with the performance-enhancing potential of gene doping, WADA uses two additional arguments for prohibiting gene doping. The first is the possible harm of gene doping for athletes. Second, is the violation of the value of fair play and the spirit of sport. (51)

As for gene therapy, every known gene can be used for gene doping. Currently, only about 500 genes in the human genome are used in existing drugs, thus a significant number of the remaining genes could bring new options for doping. (31) At least 100 genes are already linked to athletic performance and the number is increasing every year. (12,52) Although not all of these genes can be considered to be potential gene doping candidates, the increasing number of genes used in medications raises expectations for the potential advantages of gene doping.

The great benefits expected from gene doping make it likely that actual misuse is close at hand. As illustrated by the BALCO-affair, among other incidents, athletes are known to take more risks than average people. (53) BALCO was an American-based company that officially advised numerous world-class athletes on nutrition, but secretly instigated cooperation between chemists, trainers, and athletes to purposely evade doping controls with new and undetectable doping substances. The fear is that athletes might not wait for gene therapy to be fully developed and tested before misusing it. (4,9,12,19,37,53)

5.2.4.4 Gene doping targets.

Depending on the desired effect and the type of sport and athlete, gene doping might enhance performance. Athletes who compete in endurance sports, like marathons and long distance swimming, may look to gene therapy to boost their oxygen supply or delay the sense of fatigue. Sprinters and weight lifters, who mainly need power, may consider gene therapy to increase muscle mass or improve their injury recovery time. (4,13,36) Boxers would appear to be most interested in improved pain tolerance from gene doping. (12)

5.2.5 Properties, targets and current status of protein drugs and gene doping

5.2.5.1 Erythropoietin (EPO)

EPO increases oxygen supply to muscles, thereby increasing an athlete's endurance and performance. (10,13,36,54) EPO is a hormone with 165 amino acids produced mainly in the renal cortex and its production is quickly induced by hypoxia. (3,44,55-57,57) After being released, EPO binds to the erythropoietin receptor (EPOR) stimulating erythropoiesis (the production of new erythrocytes) (40,41,54,58), which increases the number of haemoglobin carrying erythrocytes in the blood. Haemoglobin binds to oxygen with high affinity; although this affinity is reduced by heat or high carbon dioxide concentrations—conditions found in active muscle tissue. (54) Thus, EPO increases the oxygen supply for muscle tissue and muscles can work longer before they build up lactic acid. (56) The result is that maximal oxygen uptake in muscles is increased, which increases endurance. (59)

Recombinant human erythropoietin (rHuEPO) was introduced in 1988 as the protein drug epoetin-alpha in Europe (and in 1989 in the US). It is used to treat anaemia caused by kidney disease, cancer or HIV, or for blood loss following surgery or trauma. Instead of giving a patient donor blood to increase erythrocytes, the patient is injected with EPO to stimulate erythropoiesis. (5,44,54-56) Although studies have shown the stimulation of steroidogenesis in Leydig cells by EPO leads to infertility over time, (44) EPO has also been found to have neuroprotective properties. (40-42,44)

The first documented illicit use of rHuEPO was in the 1989 Tour de France, (60) and more cases have since been documented. (61) Since EPO stimulates erythropoiesis, it increases the viscosity of the blood, thus raising the risk of microcirculation blockage, heart failure and strokes, (19) which makes overdosing and overexpression a risk. (3,9,19) EPO has been prohibited by the IOC since 1990 and is currently on the WADA prohibited list. (50,59,62)

Gene therapy with EPO was first tested in macaques in the late 1990s and was shown to double the number of red blood cells in ten weeks, which increased aerobic capacity and performance. Unfortunately, EPO also made the macaques' blood rather viscous, although the macaques did not go into cardiac arrest and survived.

(11,13,36) The macaques also had autoimmune reactions against EPO causing anaemia. (19,32) In a follow-up study, regulated gene expression allowed safe production of EPO for at least six years. (32) Furthermore, ex-vivo gene therapy has been performed in mice causing expression of functional EPO. (24)

It has already been shown that EPO-screening of urine samples, as currently used in WADA doping controls, can identify EPO genetic therapy. (63,64) Since muscle tissue produces EPO with posttranslational modifications that differ from EPO produced by the kidneys, illegal use could be detected using isoelectric focusing (a technique using differences in pH-dependent electric charges). (63)

On the basis of the promising animal studies, Biomedica, a British company in Oxford, developed Repoxygen to be used in the treatment of cancer, diabetic neuropathy and Parkinson's disease. Repoxygen is a viral vector containing the EPO-gene and a hypoxia-response element used to treat anaemia. However, due to safety problems in in-vivo testing—erythrocytosis, thrombosis and ischemia, and immune reactions—Repoxygen has not been clinically tested to date. (40-42) Despite the rather problematic safety profile of Repoxygen, a 2006 incident raised fear of abuse when a German track coach was accused of supplying Repoxygen to his athletes. However, the only evidence was e-mail correspondence with a Dutch general practitioner about the issue. (7,9)

In conclusion, the potential benefits and experience with EPO gene doping are quite reasonable relative to other gene-doping candidates discussed below. Although in its infancy, given the availability of an EPO gene-therapy product, it is the most likely protein to be used for gene doping. However, the availability of a broad range of conventional EPO-products and the likelihood that the current urine EPO-detection test would identify this type of gene-doping rather easily speak against the use of EPO gene doping in the sport's arena.

5.2.5.2 Insulin-like growth factor (IGF)

Although increased endurance offers major benefits for athletes like long-distance runners, EPO offers limited benefits for athletes for whom power is essential (e.g. weight lifters). For this class of athletes, IGF may be more useful since it enhances muscle growth and performance. Medical researchers currently focus on developing methods to stimulate

the endogenous production of IGF to prevent muscle loss due to a range of conditions such as degenerative muscle conditions, cancer, HIV, or aging. (5,36,65,66)

IGF-I is a polypeptide of 7.5 kDa, structurally related to insulin and produced as a result of hypothalamus-pituitary-liver axis activation. The hypothalamus produces growth-hormone- releasing hormone (GNRH), which stimulates the pituitary to release growth hormone (GH) thus stimulating the liver to produce IGF-I. (44,65,67-70) IGF stimulates muscle repair and muscle mass hypertrophy after damage, for example, from overload or stress. (13,19,44,66) Increased expression of IGF leads to increased muscle power and mass making IGF a potential target for doping. (11,66,71,72)

The effects of IGF-I on muscle growth have not been tested on humans, but in IGF-I deficient patients, insulin resistance, growth disorders and cardiovascular illnesses have all been documented. (73) Transgenic mice have been used to test the effects of IGF-I. They showed 20-50% larger muscle mass than regular mice and no age-induced muscle degradation. (36) The lifespan of these mice was decreased by 50%, possibly due to lower levels of antioxidative molecules, or cardiac hypertrophy. (73)

Although IGF-I is on the WADA prohibited list, (50) it is available on the internet (65) and anecdotal evidence proves IGF-I abuse. (74) The clear benefits of IGF—muscle growth and endurance—are desirable in many sports. The local effect of IGF allows for selective muscle growth; however, it is not expected to be one of the first targets for gene doping. The health risks of IGF gene doping, in particular, the possible clinical consequences of IGF-overexpression such as cancer and cardiac hypertrophy, are significant. (11,44)

5.2.5.3 Growth hormone (GH)

Instead of applying IGF based gene doping directly, it is possible to increase the production of IGF indirectly by aiming gene doping at the endogenous production of GH that is significantly more accessible. (74)

GH is mainly produced by the anterior pituitary gland. (75,76) The pulsatile regulation of the various GH isoforms differs in men and women and is controlled by the GH-releasing hormone, which fluctuates with sleep, exercise, hypoglycaemia, age, gender, amino-acid availability

and low levels of IGF-I. (44,70,74,75) The effects of GH are regulated by GH-binding proteins. (70)

Since GH increases muscle strength, (75) it could be used to increase athletic performances in sports where strength is important. (23,75) In endurance sports when energy is scarce, GH promotes the use of lipids as fuel to conserve protein storage. (74,76)

Despite a 1989 ban by the International Olympic Committee on GH, there is evidence of GH being used as doping. (19,51,67,74,75,77) A recent survey of 10th grade boys in the United States showed that 5% had taken GH and 1.2% of college athletes admitted to have used GH in the last year. (74,77) GH-gene therapy tests in mice, rabbits, sheep and pigs have been performed with varying results. (70,78,79) In GH-deficient mice, a 48% growth in the injected quadriceps was found after 60 days. (79) The main concerns for GH use are the lack of control in expression and disruption of functional genes. (70,78) No results of gene therapy with GH in humans have been published to date. Since the results of animal studies are far from convincing and the effects of GH are less targeted than IGF and other proteins, GH is not likely to be used as a target for gene doping. (74)

5.2.5.4 Myostatin

In 2004, a German boy born with muscular thighs and strong upper arms was diagnosed with a myostatin gene deficiency. As a result, the anti-doping community's attention was then directed towards the effects of myostatin blocking. (5,13,80,81) In cows, a myostatin mutation leads to downregulation of myostatin, which increases muscle growth. ''Double muscled cattle'' or ''Belgian Blue cattle'' present with significantly more muscle mass than ordinary cattle. (12,36) These two examples made it clear that myostatin inhibition is yet another way to increase muscle mass, but it is more specific than the use of IGF or GH. As such, myostatin inhibitors are of interest to athletes who need muscles rather than speed; however, myostatin inhibitors are on the WADA prohibited list. (5,50,81) Despite the risks of inhibiting myostatin, which include reduced cardiac and respiratory functioning, the inhibitors can be purchased on the internet. (81,82)

It is thought that myostatin is involved in sarcopenia (age-related muscle loss), although how this occurs exactly is unclear. Some tests have found increased myostatin protein and mRNA expression in aged human

and rats; others find no difference. (83) Myostatin is overexpressed in muscle atrophy when there is immobilization, HIV infection, sepsis, burn or glucocorticoid excess, or specific skeletal muscle degeneration diseases. (83-85) These findings may lead to a new treatment for muscle atrophy using gene therapy to inhibit myostatin. (85,86) Myostatin is underexpressed in Duchenne and Becker muscular dystrophy, probably as an adaptive response to increased muscle growth. (83,84) Myostatin overexpression can induce cachexia and increased levels are associated with obesity and diabetes. (80,84)

Since the actions of myostatin inhibit muscle growth, blocking myostatin is a potential doping target. (84,87) Various in-vivo methods of inhibiting myostatin are available, such as:

- using the myostatin propeptide, which binds to myostatin to prevent it from having an effect. (85,87,88) Though wild-type myostatin propeptide is unstable in-vivo, it can be altered to extend stability. (84)
- using neutralizing antibodies. (44,75,81,82,84,85,87,88) Research in mice showed less sarcopenia-related muscle loss when antibodies were injected. (89)
- applying follistatin in animal gene therapy studies to inhibit myostatin. (13,37,84-86) Follistatin is a glycoprotein that binds to myostatin preventing myostatin from binding to its receptor. (80)
- stimulating overexpression of a gene coding for a myostatin protein without its cleavage site to inhibit the production of myostatin. (81,90)

Gene therapy to inhibit myostatin is usually based on the Adeno-associated virus (AAV) vector technology, since muscle cells are one of the natural hosts for AAVs. (85,86) There is long experience with all above-presented forms of myostatin gene therapy (except antibodies) in animals; no clinical tests have been performed on humans. (80,82,84,86-88,91,92) Athletes might be tempted to use a myostatin-inhibiting form of gene doping. The effects of myostatin are significant, but the lack of experience and the poor controllability of the various methods of myostatin blocking make it hard to say whether it is already being misused.

5.2.5.5 Vascular endothelial growth factor (VEGF)

Increasing blood flow through a muscle postpones fatigue. One protein regulating muscle blood flow is VEGF or VEGF-A(93,94), also known as the vascular permeability factor. (95) Autocrine VEGF released by endothelial cells, regulates vessel homeostasis by acting as a survival factor for endothelial cells. Paracrine VEGF produced by any hypoxic cell stimulates vessel branching. (5,23,44,95-98) New capillary branches need additional hormones to become fully grown stable vessels. (98) When VEGF reaches high levels in blood vessels, the blood vessel responds with increased permeability and vasodilatation. (94,99) Since VEGF increases neovascularisation of ischemic tissue, it might help patients with heart diseases. On the other hand, high dosages of VEGF can cause vessel leakage and abnormalities as well as tumour tissue growth. (95,97,100)

Gene therapy targeting VEGF-mediated angiogenesis has been tested in mice, rats, rabbits and dogs with generally positive results. (38,94,96) In a 10-year follow-up study on humans, VEGF gene therapy was found to be safe;(93,100) therefore, VEGF might be of interest to athletes combating exhaustion. Since VEGF increases blood perfusion in muscles, heart, liver and lungs, it is likely to increase endurance. (3,23,38,44) However, the risks of VEGF use mentioned above remain unmeasured and uncontrolled. (97,101)

Controlling gene expression by adding a hypoxia-response element, e.g. EPO, might make VEGF safer. (97) VEGF is a likely candidate for gene doping;(10) however, an immune response against VEGF has been detected using affinity-based biosensors and this is likely to make detection possible soon. (23) VEGF is on the WADA-prohibited list. (50)

5.2.5.6 Fibroblast growth factor (FGF)

VEGF production is also modulated by a specific fibroblast growth factor (FGF2). FGF2 works partially in a synergistic manner with VEGF, producing some of the same intracellular effects. VEGF induces FGF2, which vice versa can induce VEGF expression. Inhibition of either VEGF or FGF2 shuts down angiogenesis. (102)

Fibroblast growth factors (FGFs) have multiple functions, some of which could be used in doping and these are discussed here. The family of FGFs includes 22 growth factors, produced by a variety of cell types. (99,102-106) The angiogenic effects of FGFs play an important role in

muscle repair following exercise through the revascularization process during muscle regeneration. (103) The modern clinical application of the principle of angiogenesis can be divided into two main areas: anti-angiogenic therapies and pro-angiogenic therapies. Whereas anti-angiogenic therapies are being employed to fight cancer and malignancies, which require an abundance of oxygen and nutrients to proliferate, pro-angiogenic therapies are being explored as options to treat cardiovascular diseases. One of the first applications of pro-angiogenic methods in humans was the use of FGF-1 for the treatment of coronary artery disease. (102,105) Clinical research in therapeutic angiogenesis is ongoing for a variety of atherosclerotic diseases, like coronary heart disease, peripheral arterial disease or wound healing disorders. (99,102,105) The risks of exogenous FGF include the possibility of increasing blood supply for tumours, or stimulating pathogenic heart remodelling. (105)

Adenovirus vectors and plasmids containing genes for FGF2 and FGF6 have been tested in human skeletal muscles and significantly increased muscle repair. (103) Phase II studies showed proteinuria as an effect of abnormal capillary network formation. (107) Most studies combine FGF gene therapy with IGF, PDGF or VEGF. (69,99,99,104,107-109) The synergistic effects of FGF with those proteins have been shown, but FGF alone has not been proven to be clinically effective. (105,109)

Most interesting for athletes is that FGFs increase muscle regeneration and neovascularisation. A combination of FGFs is most promising, especially for athletes recovering from injury and exercise. (103) It is most likely that the first use of FGF-based gene doping will be in combination with another protein. All FGFs and FGF-based gene doping are prohibited by WADA. (50)

5.2.5.7 Endorphin and Enkephalin

A completely disparate approach for improving athletic achievements is diminishing the sensation of pain. This would specifically allow combat-sport competitors to achieve higher goals. For athletes in general, numbing the sensation of extreme exhaustion is beneficial; thus, analgesics are the most frequently used therapeutic class of drugs. (10,44,51) Most analgesics are permitted by WADA, but opiates are prohibited as they have addictive properties that can lead to abuse.

Chronic pain affects a large part of the general population and gene therapy with an endorphin or enkephalin may present a promising new approach for treatment. (29,35,101,110) Endorphins and enkephalins delay fatigue and increase endurance. (13) During exercise they diminish lactic acid-related pain and pain caused by earlier injuries. (13,51) Multiple gene therapy studies aimed at combating pain have been conducted, generally with positive results. (29) Gene therapy allows for local and specific treatment of pain, with few side effects and a low risk for abuse. (35) Since HSV targets neurons specifically, this is the virus generally used for gene therapy for pain. (110) Clinical trials using endorphin and enkephalin in HSV vectors are being performed in humans, but so far are restricted to cancer-induced pain. (51,110,111)

The pain-reducing effects of both endorphin and enkephalin seem useful for athletes and early tests in humans are in progress. Given the fact that the brain is targeted, it may be difficult to detect endorphin or enkephalin gene doping in only blood or urine. It should be clear though, even for those without a biomedical education, that experimental medicines acting only on a partially understood brain system, pose a serious risk. Given the ambiguous doping qualities of endorphins and enkephalins it is rather unlikely they are being used today for this purpose.

5.2.5.8 α Actinin 3

In 2003, the association between athletic performance and α actinin 3 (ACTN3) genotype (instead of ACTN2) was demonstrated. (5) ACTN3 is mainly produced by skeletal muscle, (112) while actin and myosin are responsible for muscle contraction. (51) ACTN3 binds sarcomeres at the Z-lines. Though it was long thought that ACNT3 was only important for muscle structure, it is now clear that it is also important for muscle metabolism. (112) ACTN3 deficiency does not cause muscle disease, but rather it impairs power performance by shifting the characteristics of fast-type muscles to slow-type muscles. (112) When there is a deficiency of ACTN3, part of the action of ACTN3 is taken over by ACTN2. (36,51,112) ACTN3 expression increases strength (although androgens have a stronger effect than ACTN3(113)), while ACTN2 expression increases endurance. (51,113) No exogenous forms of ACTN3 or substances that influence ACTN3 transcription are known.

Sixteen per cent of humans worldwide have a polymorphism in both their ACTN3 genes that causes a deficiency, and in European and

Asian populations this can be up to 50%. (112,113) It has been shown that female sprinters have a higher frequency of a functioning ACTN3 gene than the average population. (36) Since lacking the gene does not cause disease, it is not a lucrative topic of research. No trial with an ACTN3 gene therapy product has been published, although there are a few animal studies on knock-out mice. Mice missing the ACTN3 gene weigh less than wild-type mice and have smaller muscles and less strength. On the other hand, they were able to run 33% longer than wild-type mice and recovered faster from fatigue. (112,113) If translated to the sports arena, this indicates that increasing ACTN3 copies may be used in order to dope sprinters and diminishing ACTN3 copies as a means to stimulate endurance in marathon runners. (51) Although both the risk of abuse and the chance of being caught would be small, no gene therapy products for ACTN3 have been tested, not even in animals. This means that currently ACTN3 is an unlikely candidate for gene doping purposes.

5.2.5.9 Peroxisome proliferator-activated receptor-delta

Another genetic predisposition for achievement in the elite sporting world is the gene coding for peroxisome proliferator-activated receptor-delta (PPARδ). PPARδ –also known as PPARβ or NR1C2– is a protein for regulating the oxidation of fatty acids. (114-116) PPARδ also increases mitochondrial activity and muscular glucose uptake. (113) Overexpression of PPARδ decreases the accumulation of triglycerides in muscle cells and increases the oxidative capacity in muscle fibres. (114,117) This results in increased endurance and an enhanced response to endurance exercise. (13,19,113,115,117) Both endurance and power training strongly increase the production of PPARδ. (115,118-121) Elite athletes have more PPARδ mRNA and protein than the general population. (19,51)

PPARδ agonists could be used for doping purposes and products claiming to boost performance with PPARδ are already for sale on the internet. (115) MBX-8025, GW742 and GW1516 (also known as GW501516) are ligands for PPARδ. They are being used in studies with patients who are obese or have diabetes mellitus type II or atherosclerosis. (23,114,117) GW1516 reduced the LDL and triglyceride plasma concentration, and increased fatty acid oxidation. (114) GW1516 has passed phase II and phase IV clinical trials for dyslipidemia (122) and is detectable with mass spectrometry up to four days after intake. (122)

Although GW1516 may be abused, the abuse might not go unpunished. Anticipating possible abuse, the WADA put PPARδ agonist GW1516 and PPARδ-AMP-activated protein kinase on the doping list in 2009. (13,50,122)

PPARδ could also be targeted with gene therapy. PPARδ has been delivered to various cell types using an adenoviral vector; however, effectiveness differed according to cell type. (116) Gene doping using PPARδ is unlikely to be used soon, since it has only been tested in cells and not in-vivo.

5.2.5.10 Cytosolic phosphoenolpyruvate carboxykinase (PEPCK-C)

An even stronger effect on endurance than PPARδ is found with PEPCK-C. In one particular study, wild-type mice were exhausted after running 0.2 km and the transgenic mice overexpressing PPARδ after 1.5 km; but the transgenic mice overexpressing PEPCK-C- ran for more than 4.9 km. (9)

PEPCK-C regulates glyconeogenesis in the liver and kidney, and glyceroneogenesis in the liver and adipose tissue. (9,123) Overexpression of PEPCK-C leads to hyperglycemia. (124) On a normal diet, PEPCK-C overexpression induces insulin sensitivity; on a fat-rich diet it causes insulin resistance. (125) Despite all the evidence of PEPCK-C's importance in gluconeogenesis, (and, thus, diabetes), no gene therapy product targeting this protein has been investigated. (123) However, silencing vectors and decreasing PEPCK-C levels did prove to be effective in diabetic animals. (123) Its role in skeletal muscles is not clear yet, but it is hypothesized that an increase in triglycerides leads to improved athletic performance. (9,13)

There are two forms of PEPCK-C: one functions in the mitochondria and one in cytosol. (9) The PEPCK-C in cytosol is the most relevant for athletes, since the effects in trials with mice demonstrated convincing benefits for endurance. (9) No specific PEPCK-C stimulating agent is yet known. (50) Thiazolidinediones and glucocorticosteroids stimulate PEPCK-C production, but not specifically. (126) Glucocorticosteroids are on the WADA prohibited-list; however, to date, no gene therapy product aimed at PEPCK-C specifically is known. (50) Since the expression of the PEPCK-C gene would be tissue-specific, detection would be nearly impossible; (9) thus, in addition to the

significant effect of PEPCK-C, despite the lack of experience, PEPCK-C is a likely doping target.

Table 2. This table summarizes the likelihood of each protein being used for gene doping in a scale ranging from ++ (very high) to - - (very low). For each protein, the average of the possible methods is considered (i.e., the multiple options for inhibiting myostatin taken on average, just like the various FGF's). If it is useful for athletes to abuse a protein, then it is marked in the first column as potential benefits. The experience in gene therapy-column has marks concerning the experience of gene therapy in humans. When fully developed, some proteins remain more dangerous than others and this is noted in the third column, risk control. The chances of getting away with the illegitimate use of a gene-doping product with a specific protein are given in the column chance of undetected use. The last column indicates the likelihood of present abuse.

Protein	Potential benefits	Experience in gene therapy	Risk control	Chance of undetected use	Likelihood of abuse
EPO	++	+	±	- -	++
IGF	++	-	-	±	-
GH	±	-	-	-	-
Myostatin inhibition	+	±	±	±	±
VEGF	+	++	-	-	++
FGFs	±	±	±	±	±
Endorphin and Enkephalin	+	±	- -	++	-
ACTN3	+	- -	+	++	- -
PPARδ	++	-	+	±	+
PEPCK-C	++	- -	±	++	+

Table 3. This table classifies each reviewed protein as either functional for improving endurance, strength or pain tolerance.

Aimed enhancement	Target gene
Endurance	EPO, IGF-I, GH, VEGF, FGF1, FGF2, FGF4, ACTN2, PPARδ, PEPCK-C, IGF-I
Strength	FGF6, FGF2, IGF-I, GH, Myostatin, ACTN3
Pain tolerance	Endorphin, Enkephalin

5.2.6 Detection

Detection of gene doping is significantly more difficult than detection of doping with pharmaceuticals. This might make gene doping more attractive for athletes considering cheating. (23,44,127,128) Currently, no specific test to detect gene doping has been approved by the WADA or used by a WADA-accredited laboratory. (5,10,23,129,130)

Any detection method would have to comply with at least the following requirements. First, the doping detection method must be adequately selective to detect cheating athletes. Second, it should be accessible and easy to use on a large scale, while remaining reliable. Finally, it should be fast, as convicting an athlete years after the crime is not desirable (although legally possible up to eight years after a doping violation has occurred). (13,128)

As stated earlier, athletes who engage in doping generally use pharmaceuticals to improve their performances, so the first gene dopers are likely to have had early access to gene therapy products. Detection efforts to identify gene doping by athletes should initially explore the current uses of known gene therapy for disease treatment. (12) Some detection methods might also help to determine the efficacy of gene therapy for disease while in development. (128)

Generally, detection methods can be divided into two groups: direct and indirect. Direct methods test for an illegal substance, or the genetic material or virus that delivered it. Indirect methods use the effect, immune response, differences in expression, or metabolic changes for detection. (12,13,44) The direct detection of an illegal substance is preferred over indirect testing for legal reasons, but unfortunately, illegal substances are metabolized or cleared too quickly to be detected, in general. (12,23,75)

5.2.6.1 Direct methods

5.2.6.1.1 Plasma levels

Measuring the plasma levels of a protein would not be an accurate method for detecting gene doping. Some endogenous mechanisms to control expression prevent high plasma levels and the plasma levels of some proteins are too low to detect. (19). Also fluctuation in physiologic levels of a protein complicates this method. Measuring various isoforms of a protein would be helpful. When an exogenous protein inhibits production of the endogenous variant, a difference in the isoform ratio

would be detected; thus, detection would be possible for gene doping strategies targeting EPO and GH. (75)

5.2.6.1.2 Biopsy.

A biopsy of an infected area would provide a sample in which the virus or the exogenous gene might be detectable in an athlete. Gene transfer has been shown to be detectable in a biopsy for up to a decade. (48) Since knowledge about the injection site is required, and biopsies are generally considered to be too invasive, methods using only blood, urine, serum, hair, saliva or a combination are needed. (6,23,36,75,128,130)

Virus. The presence of a virus might be detectable in the bloodstream, so blood samples could be tested with PCR to detect DNA or RNA or with other methods to test for viral proteins. (12,23,30) The difficulty with this technique is timing; the persistence of viruses varies from hours to months. (23,44) Testing for a virus in urine (persistence over several weeks) or saliva (persistence over several days) might be better. (23) The downside of this approach is a possible false positive, e.g., an athlete who is infected with a normal virus. (12,36,75)

5.2.6.1.3 Introns.

The genetic material commonly used in gene doping is complementary DNA (cDNA), which lacks introns; therefore, it can be discriminated from genomic DNA with PCR. (12,37,44,127,129) In mice injected intramuscularly with AAV-mediated gene therapy, PCR allowed detection in the blood for several weeks, (37) though in another test it was undetectable in blood after half an hour. (130) However, PCR is less useful for detecting doping using genes with introns because it presents problems with alternative splicing and efficacy. (12,127) In addition, it is conceivable that once a PCR detection method is introduced, gene-doping products based on genomic DNA will become available quite soon.

Posttranslational modification. Since each cell type differs in posttranslational modification, endogenous proteins would be distinguishable from the ones produced by gene doping. This is what caused the autoimmune reaction against EPO in macaques, resulting in anaemia. (19,32) Detection would be possible with isoelectric focussing. (6,10,12,19,23,44) The method is useful until viruses target cells more specifically, or specific promoter regions are developed. However, it is possible that these target cells or promoter regions might also be detectable in the future. (12,23)

Barcoding. Genetically modified agricultural products have a genetic barcode, to help with identification. This could be done for gene therapy too, which would make the gene therapy products detectable with PCR. This approach requires global coordination in the pharmaceutical industry, which in the past has been proven to be difficult to achieve, and is likely to become practically irrelevant once gene-doping products are produced without barcodes. (6,12,51,129) Creating barcodes for identification could stimulate rogue laboratory production practices that eliminate barcodes.

5.2.6.2 Indirect methods

5.2.6.2.1 Immune reaction.

Every virus induces a specific immune response in the host. (6,12,23,44) Plasmid vectors or the produced proteins can induce immune responses that can be detected and distinguished from common immune responses. (23) However, distinguishing common virus reactions from immune reactions remains a problematic issue.

5.2.6.2.2 Proteomic changes.

Use of gene doping will probably change the transcription of other proteins as well. By tracking selected protein levels and gene transcription rates in a biomedical passport, dopers can be caught. (4,6,12,13,19,44,51,54,129) One risk of this method is the chance of a false-positive or false-negative, since changes in training or injuries can also induce changes in metabolism. Most research on possible detection methods of gene doping uses this approach, also because this approach is potentially quite useful to determine the efficacy of gene therapy trials; but the validity is as yet unproven. (37,47,47,54)

In conclusion, gene-doping detection is difficult, but with new techniques it might eventually be possible. (10,12,23,47) False positives are a nightmare to every anti-doping professional, so it is important to validate detection methods before applying them. It is likely that once a test has been developed, it will not be made public until it can be used for anti-doping, just as previously done for detecting hydroxyethyl starch or homologous blood transfusions among other prohibited substances in sport.

5.2.7 Animal Use of Gene Doping

Since gene doping can increase performance, it is likely to be used in animal competitions as well. If money can be earned by betting or trading

with superior animals, gene doping would be lucrative. If the achievements of horses, dogs, camels or pigeons could be improved, then it is quite possible that gene doping will be tried on these animals before human applications.

5.2.8 Conclusion

Before each post-2000 Olympic Game, the media have predicted gene doping. So far, all predictions have been proven wrong, based on the information that is currently available.

Even though gene doping can be done with the lab skills of an undergraduate student, this does not mean that it is actually being applied in athletics. In addition, most currently available gene-therapy products still show a rather low efficiency of gene transfer and have side effects, some of which can be quite serious. (7) Thus, the question remains: is it likely gene doping is already being used and will be used at the 2016 Rio de Janeiro Olympics?

This is still a realistic option, despite all the efforts of anti-doping professionals. All proteins reviewed in this survey have the potential of performance-enhancing effects in sports and can be targeted with gene therapy. Some have already been tested in gene-therapy animal experiments and in clinical trials, like EPO and VEGF. These proteins are potentially the most likely candidates to be misused, but they are also the ones with the highest risk of detection and the ones that can be applied more cost-effectively by conventional means. Other proteins have only recently been selected for gene therapy research purposes. We have identified PPARδ and PEPCK-C as having high potential for abuse. But we expect that for efficiency reasons, there will be a preference for inserting gene target combinations rather than single gene doping products. This will also further complicate detection.

However, it is still fair to say that there is no clear proof that gene doping is already practised in major sporting competitions. Given the current niche status of gene therapy, it is not realistic to estimate the time period when gene doping will enter athletics. The interest is obviously there, and historically it is known that the determined cheat will try almost anything to boost their performance, regardless of the risks involved. Although gene doping is still largely theoretical, its implications for sports, health, and ethics are significant and require further study.

5.2.8.1 Author's contributions:

All authors were involved in the design of the systematic review, drafting of earlier versions of the manuscript, and providing final approval for submission. TvdG was responsible for the collection, analysis, and interpretation of the systematic review data, as well as drafting, and revising the manuscript. TP was responsible for the analysis and interpretation of the data, drafting supervision, and the revisions. OdH and HH provided support in all phases of the research process.

5.2.8.2 Conflict of interest

None declared.

5.2.8.3 Source of funding

Regular institutional funding for all authors.

5.2.8.4 Further reading

Gene Therapy Clinical Trials Worldwide--The Journal of Gene Medicine Clinical Trial site. www.abedia.com/wiley/index.html

5.2.8.5 Acknowledgements

The authors would like to thank Julia Challinor for her English manuscript correction services and the reviewers for their critical but constructive comments.

5.2.9 References

1. Friedmann T, Koss JO. Gene transfer and athletics- an impending problem. Mol Ther 2001;3:819-820.
2. KATERSKY A. 'Gene Doping' a New Threat to Olympic Fairness. ABC news 2006;feb 9.
3. Unal M, Ozer Unal D. Gene doping in sports. Sports Med 2004;34:357-362.
4. Filipp F. Is science killing sport? Gene therapy and its possible abuse in doping. EMBO Rep 2007;8:433-435.
5. Ostrander EA, Huson HJ, Ostrander GK. Genetics of athletic performance. Annu Rev Genomics Hum Genet 2009;10:407-429.
6 .Azzazy HM, Mansour MM, Christenson RH. Doping in the recombinant era: strategies and counterstrategies. Clin Biochem 2005;38:959-965.
7. Friedmann T. How close are we to gene doping? Hastings Cent Rep 2010;40:20-22.
8. WADA M. 2010 Adverse Analytical Findings and Atypical Findings Reported by Accredited Laboratories. 2011.
9. Azzazy HM, Mansour MM, Christenson RH. Gene doping: of mice and men. Clin Biochem 2009;42:435-441.
10. Haisma HJ, de Hon O. Gene doping. Int J Sports Med 2006;27:257-266.

11. Harridge SD, Velloso CP. IGF-I and GH: potential use in gene doping. Growth Horm IGF Res 2009;19:378-382.
12. Mansour MM, Azzazy HM. The hunt for gene dopers. Drug Test Anal 2009;1:311-322.
13. McKanna TA, Toriello HV. Gene doping: the hype and the harm. Pediatr Clin North Am 2010;57:719-727.
14. Kay MA. State-of-the-art gene-based therapies: the road ahead. Nat Rev Genet 2011;12:316-328.
15. Pearson S, Jia H, Kandachi K. China approves first gene therapy. Nat Biotechnol 2004;22:3-4.
16 European Medicines Agency Press Office. European Medicines Agency recommends first gene therapy for approval. 2012.
17. Sheridan C. Gene therapy finds its niche. Nat Biotechnol 2011;29:121-128.
18. Hojman P, Gissel H, Gehl J. Sensitive and precise regulation of haemoglobin after gene transfer of erythropoietin to muscle tissue using electroporation. Gene Ther 2007;14:950-959.
19. Wells DJ. Gene doping: the hype and the reality. Br J Pharmacol 2008;154:623-631.
20. Moelling K. Naked DNA--the poor man's gene therapy? Gene Ther 1998;5:573-574.
21. Witzenbichler B, Mahfoudi A, Soubrier F, et al. Intramuscular gene transfer of fibroblast growth factor-1 using improved pCOR plasmid design stimulates collateral formation in a rabbit ischemic hindlimb model. J Mol Med (Berl) 2006;84:491-502.
22. Mingozzi F, High KA. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. Nat Rev Genet 2011;12:341-355.
23. Minunni M, Scarano S, Mascini M. Affinity-based biosensors as promising tools for gene doping detection. Trends Biotechnol 2008;26:236-243.
24. Lin RZ, Dreyzin A, Aamodt K, et al. Induction of erythropoiesis using human vascular networks genetically engineered for controlled erythropoietin release. Blood 2011;118:5420-5428.
25. Kim MH, Cho GW, Huh YM, et al. Transduction of human EPO into human bone marrow mesenchymal stromal cells synergistically enhances cell-protective and migratory effects. Mol Biol (Mosk) 2010;44:656-663.
26. Odom GL, Gregorevic P, Chamberlain JS. Viral-mediated gene therapy for the muscular dystrophies: successes, limitations and recent advances. Biochim Biophys Acta 2007;1772:243-262.
27. Thaci B, Ulasov IV, Wainwright DA, et al. The challenge for gene therapy: innate immune response to adenoviruses. Oncotarget 2011;2:113-121.
28. Hao S, Mata M, Fink DJ. Viral vector-based gene transfer for treatment of chronic pain. Int Anesthesiol Clin 2007;45:59-71.

29. Cope DK, Lariviere WR. Gene therapy and chronic pain. ScientificWorldJournal 2006;6:1066-1074.
30. Eslami MH, Gangadharan SP, Sui X, et al. Gene delivery to in situ veins: differential effects of adenovirus and adeno-associated viral vectors. J Vasc Surg 2000;31:1149-1159.
31. Black JL,3rd. Genome projects and gene therapy: gateways to next generation biological weapons. Mil Med 2003;168:864-871.
32. Rivera VM, Gao GP, Grant RL, et al. Long-term pharmacologically regulated expression of erythropoietin in primates following AAV-mediated gene transfer. Blood 2005;105:1424-1430.
33. Yi Y, Noh MJ, Lee KH. Current advances in retroviral gene therapy. Curr Gene Ther 2011;11:218-228.
34. Guse K, Cerullo V, Hemminki A. Oncolytic vaccinia virus for the treatment of cancer. Expert Opin Biol Ther 2011;11:595-608.
35. Jain KK. Gene therapy for pain. Expert Opin Biol Ther 2008;8:1855-1866.
36. Sweeney HL. Gene doping. Sci Am 2004;291:62-69.
37. Beiter T, Zimmermann M, Fragasso A, et al. Direct and long-term detection of gene doping in conventional blood samples. Gene Ther 2011;18:225-231.
38. Karvinen H, Pasanen E, Rissanen TT, et al. Long-term VEGF-A expression promotes aberrant angiogenesis and fibrosis in skeletal muscle. Gene Ther 2011.
39. Mata M, Hao S, Fink DJ. Applications of gene therapy to the treatment of chronic pain. Curr Gene Ther 2008;8:42-48.
40. Puskovic V, Wolfe D, Wechuck J, et al. HSV-mediated delivery of erythropoietin restores dopaminergic function in MPTP-treated mice. Mol Ther 2006;14:710-715.
41. Wu Z, Mata M, Fink DJ. Prolonged regulatable expression of EPO from an HSV vector using the LAP2 promoter element. Gene Ther 2011.
42. Xue YQ, Ma BF, Zhao LR, et al. AAV9-mediated erythropoietin gene delivery into the brain protects nigral dopaminergic neurons in a rat model of Parkinson's disease. Gene Ther 2010;17:83-94.
43. Machelska H, Schroff M, Oswald D, et al. Peripheral non-viral MIDGE vector-driven delivery of beta-endorphin in inflammatory pain. Mol Pain 2009;5:72.
44. Oliveira RS, Collares TF, Smith KR, et al. The use of genes for performance enhancement: doping or therapy? Braz J Med Biol Res 2011;44:1194-1201.
45. Sharp NC. The human genome and sport, including epigenetics and athleticogenomics: a brief look at a rapidly changing field. J Sports Sci 2008;26:1127-1133.
46. European Society of Gene Therapy. French gene therapy group reports on the adverse event in a clinical trial of gene therapy for X-linked

severe combined immune deficiency (X-SCID). Position statement from the European Society of Gene Therapy. J Gene Med 2003;5:82-84.
47. Friedmann T, Rabin O, Frankel MS. Ethics. Gene doping and sport. Science 2010;327:647-648.
48. Buchlis G, Podsakoff GM, Radu A, et al. Factor IX expression in skeletal muscle of a severe hemophilia B patient 10 years after AAV-mediated gene transfer. Blood 2012;119:3038-3041.
49. Cavazzana-Calvo M, Thrasher A, Mavilio F. The future of gene therapy. Nature 2004;427:779-781.
50. World Anti-Doping Agency. THE 2012 PROHIBITED LIST. 2011;2012:9.
51. Gaffney GR, Parisotto R. Gene doping: a review of performance-enhancing genetics. Pediatr Clin North Am 2007;54:807-22, xii-xiii.
52. Hagberg JM, Rankinen T, Loos RJ, et al. Advances in exercise, fitness, and performance genomics in 2010. Med Sci Sports Exerc 2011;43:743-752.
53. Schneider AJ, Friedmann T. Gene doping in sports: the science and ethics of genetically modified athletes. Adv Genet 2006;51:1-110.
54. Elliott S. Erythropoiesis-stimulating agents and other methods to enhance oxygen transport. Br J Pharmacol 2008;154:529-541.
55. Szenajch J, Wcislo G, Jeong JY, et al. The role of erythropoietin and its receptor in growth, survival and therapeutic response of human tumor cells From clinic to bench - a critical review. Biochim Biophys Acta 2010;1806:82-95.
56. Tsitsimpikou C, Kouretas D, Tsarouhas K, et al. Applications and biomonitoring issues of recombinant erythropoietins for doping control. Ther Drug Monit 2011;33:3-13.
57 .Hardee ME, Arcasoy MO, Blackwell KL, et al. Erythropoietin biology in cancer. Clin Cancer Res 2006;12:332-339.
58. Ashenden MJ, Hahn AG, Martin DT, et al. A comparison of the physiological response to simulated altitude exposure and r-HuEpo administration. J Sports Sci 2001;19:831-837.
59 .Birkeland KI, Stray-Gundersen J, Hemmersbach P, et al. Effect of rhEPO administration on serum levels of sTfR and cycling performance. Med Sci Sports Exerc 2000;32:1238-1243.
60. Smeets M. Het laatste geel. Amsterdam: Nieuw Amsterdam 2009:256.
61. Lage JM, Panizo C, Masdeu J, et al. Cyclist's doping associated with cerebral sinus thrombosis. Neurology 2002;58:665.
62. McGrath JC, Cowan DA. Drugs in sport. Br J Pharmacol 2008;154:493-495.
63. Lasne F, Martin L, de Ceaurriz J, et al. "Genetic Doping" with erythropoietin cDNA in primate muscle is detectable. Mol Ther 2004;10:409-410.
64. World Anti-Doping Agency. World Anti-Doping Code. 2007:136.
65. Goldspink G, Wessner B, Bachl N. Growth factors, muscle function and doping. Curr Opin Pharmacol 2008;8:352-357.

66. Macedo A, Moriggi M, Vasso M, et al. Enhanced Athletic Performance upon Multi-Site Aav-Igf1 Gene Transfer Coincides with Massive Modification of the Muscle Proteome. Hum Gene Ther 2011.
67. Sonksen PH, Holt RI. GH & IGF Research issue on doping with growth hormone. Growth Horm IGF Res 2009;19:283-284.
68. Bosch-Marce M, Wee CD, Martinez TL, et al. Increased IGF-1 in muscle modulates the phenotype of severe SMA mice. Hum Mol Genet 2011;20:1844-1853.
69. Creaney L, Hamilton B. Growth factor delivery methods in the management of sports injuries: the state of play. Br J Sports Med 2008;42:314-320.
70. Rodriguez S, Gaunt TR, Day IN. Molecular genetics of human growth hormone, insulin-like growth factors and their pathways in common disease. Hum Genet 2007;122:1-21.
71. Annunziata M, Granata R, Ghigo E. The IGF system. Acta Diabetol 2011;48:1-9.
72. Kiuru M, Crystal RG. Progress and prospects: gene therapy for performance and appearance enhancement. Gene Ther 2008;15:329-337.
73. Belfiore A, Frasca F, Pandini G, et al. Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease. Endocr Rev 2009;30:586-623.
74. Holt RI, Sonksen PH. Growth hormone, IGF-I and insulin and their abuse in sport. Br J Pharmacol 2008;154:542-556.
75. Segura J, Gutierrez-Gallego R, Ventura R, et al. Growth hormone in sport: beyond Beijing 2008. Ther Drug Monit 2009;31:3-13.
76. Moller N, Jorgensen JO. Effects of growth hormone on glucose, lipid, and protein metabolism in human subjects. Endocr Rev 2009;30:152-177.
77. Ho KK, Nelson AE. Growth hormone in sports: detecting the doped or duped. Horm Res Paediatr 2011;76 Suppl 1:84-90.
78. Bigey P, Burgain A, Escriou V, et al. Recent progress in non viral gene delivery and therapy. Hum Gene Ther 2011;22:A7.
79. Oliveira NA, Cecchi CR, Higuti E, et al. Long-term human growth hormone expression and partial phenotypic correction by plasmid-based gene therapy in an animal model of isolated growth hormone deficiency. J Gene Med 2010;12:580-585.
80. McPherron AC. Metabolic Functions of Myostatin and Gdf11. Immunol Endocr Metab Agents Med Chem 2010;10:217-231.
81. Fedoruk MN, Rupert JL. Myostatin inhibition: a potential performance enhancement strategy? Scand J Med Sci Sports 2008;18:123-131.
82. Bachl N, Derman W, Engebretsen L, et al. Therapeutic use of growth factors in the musculoskeletal system in sports-related injuries. J Sports Med Phys Fitness 2009;49:346-357.

83. McFarlane C, Sharma M, Kambadur R. Myostatin is a procachectic growth factor during postnatal myogenesis. Curr Opin Clin Nutr Metab Care 2008;11:422-427.
84. Huang Z, Chen X, Chen D. Myostatin: a novel insight into its role in metabolism, signal pathways, and expression regulation. Cell Signal 2011;23:1441-1446.
85. Schakman O, Thissen JP. Gene therapy with anabolic growth factors to prevent muscle atrophy. Curr Opin Clin Nutr Metab Care 2006;9:207-213.
86. Rodino-Klapac LR, Haidet AM, Kota J, et al. Inhibition of myostatin with emphasis on follistatin as a therapy for muscle disease. Muscle Nerve 2009;39:283-296.
87. Morine KJ, Bish LT, Pendrak K, et al. Systemic myostatin inhibition via liver-targeted gene transfer in normal and dystrophic mice. PLoS One 2010;5:e9176.
88. Bartoli M, Poupiot J, Vulin A, et al. AAV-mediated delivery of a mutated myostatin propeptide ameliorates calpain 3 but not alpha-sarcoglycan deficiency. Gene Ther 2007;14:733-740.
89. Murphy KT, Koopman R, Naim T, et al. Antibody-directed myostatin inhibition in 21-mo-old mice reveals novel roles for myostatin signaling in skeletal muscle structure and function. FASEB J 2010;24:4433-4442.
90. Joulia-Ekaza D, Cabello G. Myostatin regulation of muscle development: molecular basis, natural mutations, physiopathological aspects. Exp Cell Res 2006;312:2401-2414.
91. Carnac G, Vernus B, Bonnieu A. Myostatin in the pathophysiology of skeletal muscle. Curr Genomics 2007;8:415-422.
92. Kota J, Handy CR, Haidet AM, et al. Follistatin gene delivery enhances muscle growth and strength in nonhuman primates. Sci Transl Med 2009;1:6ra15.
93. Lavu M, Gundewar S, Lefer DJ. Gene therapy for ischemic heart disease. J Mol Cell Cardiol 2011;50:742-750.
94. Dai J, Rabie AB. VEGF: an essential mediator of both angiogenesis and endochondral ossification. J Dent Res 2007;86:937-950.
95. Carmeliet P, Jain RK. Molecular mechanisms and clinical applications of angiogenesis. Nature 2011;473:298-307.
96. Dicks D, Saloner D, Martin A, et al. Percutaneous transendocardial VEGF gene therapy: MRI guided delivery and characterization of 3D myocardial strain. Int J Cardiol 2010;143:255-263.
97. Kim HA, Lim S, Moon HH, et al. Hypoxia-inducible vascular endothelial growth factor gene therapy using the oxygen-dependent degradation domain in myocardial ischemia. Pharm Res 2010;27:2075-2084.
98. Kolostova K, Taltynov O, Pinterova D, et al. Wound healing gene therapy: cartilage regeneration induced by vascular endothelial growth factor plasmid. Am J Otolaryngol 2011.

99. Adas G, Percem A, Adas M, et al. VEGF-A and FGF gene therapy accelerate healing of ischemic colonic anastomoses (experimental study). Int J Surg 2011.
100. Muona K, Makinen K, Hedman M, et al. 10-year safety follow-up in patients with local VEGF gene transfer to ischemic lower limb. Gene Ther 2011.
101. Liu GS, Liu LF, Lin CJ, et al. Gene transfer of pro-opiomelanocortin prohormone suppressed the growth and metastasis of melanoma: involvement of alpha-melanocyte-stimulating hormone-mediated inhibition of the nuclear factor kappaB/cyclooxygenase-2 pathway. Mol Pharmacol 2006;69:440-451.
102. Maulik N, Thirunavukkarasu M. Growth factors and cell therapy in myocardial regeneration. J Mol Cell Cardiol 2008;44:219-227.
103. Yun YR, Won JE, Jeon E, et al. Fibroblast growth factors: biology, function, and application for tissue regeneration. J Tissue Eng 2010;2010:218142.
104. Jazwa A, Kucharzewska P, Leja J, et al. Combined vascular endothelial growth factor-A and fibroblast growth factor 4 gene transfer improves wound healing in diabetic mice. Genet Vaccines Ther 2010;8:6.
105. Kapur NK, Rade JJ. Fibroblast growth factor 4 gene therapy for chronic ischemic heart disease. Trends Cardiovasc Med 2008;18:133-141.
106. Pownall ME, Isaacs HV. FGF Signalling in Vertebrate Development. 2010.
107. de Paula EV, Flores-Nascimento MC, Arruda VR, et al. Dual gene transfer of fibroblast growth factor-2 and platelet derived growth factor-BB using plasmid deoxyribonucleic acid promotes effective angiogenesis and arteriogenesis in a rodent model of hindlimb ischemia. Transl Res 2009;153:232-239.
108. Madry H, Orth P, Kaul G, et al. Acceleration of articular cartilage repair by combined gene transfer of human insulin-like growth factor I and fibroblast growth factor-2 in vivo. Arch Orthop Trauma Surg 2010;130:1311-1322.
109. Yau TM, Kim C, Li G, et al. Enhanced angiogenesis with multimodal cell-based gene therapy. Ann Thorac Surg 2007;83:1110-1119.
110. Glorioso JC, Fink DJ. Gene therapy for pain: introduction to the special issue. Gene Ther 2009;16:453-454.
111. Pohl M, Fink DJ. A new player in gene therapy for pain? Gene Ther 2008;15:953-954.
112. Berman Y, North KN. A gene for speed: the emerging role of alpha-actinin-3 in muscle metabolism. Physiology (Bethesda) 2010;25:250-259.
113. Lippi G, Longo UG, Maffulli N. Genetics and sports. Br Med Bull 2010;93:27-47.

114. Karpe F, Ehrenborg EE. PPARdelta in humans: genetic and pharmacological evidence for a significant metabolic function. Curr Opin Lipidol 2009;20:333-336.
115. Ehrenborg E, Krook A. Regulation of skeletal muscle physiology and metabolism by peroxisome proliferator-activated receptor delta. Pharmacol Rev 2009;61:373-393.
116. Nielsen R, Grontved L, Stunnenberg HG, et al. Peroxisome proliferator-activated receptor subtype- and cell-type-specific activation of genomic target genes upon adenoviral transgene delivery. Mol Cell Biol 2006;26:5698-5714.
117. Ahmetov II, Astratenkova IV, Rogozkin VA. Association of a PPARD polymorphism with human physical performance. Mol Biol (N Y) 2007;41:776-780.
118. Eynon N, Meckel Y, Alves AJ, et al. Is there an interaction between PPARD T294C and PPARGC1A Gly482Ser polymorphisms and human endurance performance? Exp Physiol 2009;94:1147-1152.
119. Grarup N, Albrechtsen A, Ek J, et al. Variation in the peroxisome proliferator-activated receptor delta gene in relation to common metabolic traits in 7,495 middle-aged white people. Diabetologia 2007;50:1201-1208.
120. Remels AH, Gosker HR, Schrauwen P, et al. Peroxisome proliferator-activated receptors: a therapeutic target in COPD? Eur Respir J 2008;31:502-508.
121. Stefan N, Thamer C, Staiger H, et al. Genetic variations in PPARD and PPARGC1A determine mitochondrial function and change in aerobic physical fitness and insulin sensitivity during lifestyle intervention. J Clin Endocrinol Metab 2007;92:1827-1833.
122. Thevis M, Moller I, Thomas A, et al. Characterization of two major urinary metabolites of the PPARdelta-agonist GW1516 and implementation of the drug in routine doping controls. Anal Bioanal Chem 2010;396:2479-2491.
123. Gomez-Valades AG, Vidal-Alabro A, Molas M, et al. Overcoming diabetes-induced hyperglycemia through inhibition of hepatic phosphoenolpyruvate carboxykinase (GTP) with RNAi. Mol Ther 2006;13:401-410.
124. Juan YC, Chang CC, Tsai WJ, et al. Pharmacological evaluation of insulin mimetic novel suppressors of PEPCK gene transcription from Paeoniae Rubra Radix. J Ethnopharmacol 2011;137:592-600.
125. Franckhauser S, Munoz S, Elias I, et al. Adipose overexpression of phosphoenolpyruvate carboxykinase leads to high susceptibility to diet-induced insulin resistance and obesity. Diabetes 2006;55:273-280.
126. Cadoudal T, Blouin JM, Collinet M, et al. Acute and selective regulation of glyceroneogenesis and cytosolic phosphoenolpyruvate

carboxykinase in adipose tissue by thiazolidinediones in type 2 diabetes. Diabetologia 2007;50:666-675.
127. Baoutina A, Coldham T, Bains GS, et al. Gene doping detection: evaluation of approach for direct detection of gene transfer using erythropoietin as a model system. Gene Ther 2010;17:1022-1032.
128. Ni W, Le Guiner C, Gernoux G, et al. Longevity of rAAV vector and plasmid DNA in blood after intramuscular injection in nonhuman primates: implications for gene doping. Gene Ther 2011;18:709-718.
129. Beiter T, Zimmermann M, Fragasso A, et al. Establishing a novel single-copy primer-internal intron-spanning PCR (spiPCR) procedure for the direct detection of gene doping. Exerc Immunol Rev 2008;14:73-85.
130. Carter A, Flueck M. A polymerase chain reaction-based methodology to detect gene doping. Eur J Appl Physiol 2011.

5.3 Evaluation

5.3.1 Strengths

This SR evaluated the most likely abused targets for gene doping. The search strategy returned relevant research that allowed us to estimate the chances of abuse for each selected target. The scoring of the proteins, which attempts to add a quantitative element to objectively score an outcome to the research question (the estimated likelihood of abuse per protein), helped in selecting the most likely targets.

The strengths of this paper are the comprehensive overview it gives of the state of the art of gene doping, using three scientific search engines. We presented a scoring framework, a systematic approach to reach a reproducible conclusion about the methods that are most likely to be used for performance enhancement with gene doping. The additional information presented, about possible vectors, the challenges in gene doping, and the detection methods, help contextualize the research question so that the reader understands the landscape and challenges.

5.3.2 Limitations

One of the main weaknesses is that the selection of the eleven selected proteins, which were used as a starting point, was based on one previous review. We did not independently assess if there were any other proteins that could have been used. This leaves a possible blind spot outside the search we performed.

Another weakness is the limitation on inclusion "until a saturation point had been reached." This made the review more feasible, as there was a lot of duplicative or irrelevant information on some of the selected proteins, but it might mean that some essential papers have been missed.

The information about vectors and mechanisms was insightful and helped to frame the landscape of gene doping, but was not a goal of the review, and was not specifically searched for. This means that we might have missed publications contradicting our sources, making our overview less reliable.

Though our scoring methodology seemed insightful, it was not described in detail, and this was not done independently by multiple authors. Though we did discuss the data in the table with the team, this

was ad-hoc and based on an already filled-in table, which creates the possibility of an anchoring bias.

5.3.3 Appropriateness of the methodology

The aim of this review was to "present the current status of gene doping and detection methods in order to answer the following pressing question: Is it likely gene doping will be used at major sporting competitions, and if so, what are the odds of effective detection?"

Using separate searches for each compound in three databases provided a clear and comprehensive overview of the status of gene doping. Limiting to only these compounds was appropriate for the goal of repeating the research done five years prior, which predicted the breakthrough of gene doping in five years. There may have been other compounds we did not look at, though, which is a major weakness of this study.

5.3.4 Lessons learned

If we were to set up this methodology again, and do it in the best possible way, we would have to set up clear criteria for the inclusion and exclusion of the proteins that are to be researched. We should also continue to include papers despite saturation, or at least define saturation with a minimum number of papers for each of the proteins before saturation can be claimed. We should improve the scoring criteria and methodology by having independent reviewers score and reconcile based on a discussion. Finally, we might include newspaper sources, as these would be more likely to give a timely view into any experiences that would not make it to the scientific literature, such as banned athletes or coaches, or inquiries that signal interest.

Whether this would have led us to a different conclusion is hard to estimate. We might have found more target proteins if we had searched more comprehensively and might have included more information which could have changed the scoring. Only time will tell if any of these proteins will indeed be used for gene doping and if the estimated probabilities are accurate.

The impact of this publication on the research field is clear. With 38 citations recorded in Google Scholar up to August 2019 and marked as "In the top 5% of all research outputs scored by Altmetric," it is clear it gained traction in the scientific community. The Dutch Anti-Doping

Authority added the publication to their list of key publications in their annual report of 2013.[179]

If we were to repeat this review to estimate the likelihood of gene doping use during the 2020 summer Olympics in Japan, we would have two options. First, we could use the same proteins, trying to validate our previous findings. Second, we could start with an open search to establish which proteins have a beneficial effect on athletic performance. Based on that, we could set up individual searches for the proteins that we selected. This second option would be more reliable, but require much more time and effort. Mainly for the first option, there is a risk that a compound that was not included in the search is used for gene doping first.

This research field does not allow for more advanced research techniques, such as meta-regression and network analyses. They require more comparable inputs, and this research field is not mature enough. In term, it might be possible to compare animal research on several of these compounds. This would require establishing the best output, either force or endurance, for each protein, so that the effect can be measured and compared.

6 Systematic review of the mechanisms and evidence behind the hypocholesterolaemic effects of HPMC, pectin and chitosan in animal trials

As the last three chapters have provided ample insights into how to find the right sources for an SR, this SR will look at how to assess the validity of the data. Unlike in chapter 5, this question allowed for relatively straightforward ways of integrating the data and calculating an effect. Therefore, we quantitively assessed the data and found relatively consistent overall effects.

But that does not automatically mean the effect is reliable. We discussed in paragraph 2.1.5 how publication bias is a significant problem for SRs, as there might be relevant data that was never published. If this data is systematically withheld for a specific reason, for example lack of effect, this can colour the results of an SR. Though trials in humans are registered, animal trials are usually not, so this landscape is particularly susceptible to publication bias.

To assess if this was the case, we developed a funnel plot (as mentioned in paragraph 1.3.9). This allowed us to assess whether there was a selection in publication, assuming the variation in results of trials would follow an expected pattern: more variation with fewer subjects, more certainty with more subjects.

6.1 Context

This systematic review looked at hypocholesterolaemic fibres in animal trials that had been identified in the literature. The research objective was to determine the state of the evidence for the cholesterol-lowering effects of three selected fibres and their mechanisms, using the most recent animal trials.

The search was performed in Pubmed, Embase and the Cochrane Library, specific for each fibre, limiting to 1 January 2000 to 2 September 2014. HPMC, chitosan and pectin were selected for review, based on: (1) the evidence of their cholesterol-lowering effect. Many studies on these fibres have been published, so information is likely to be available, which allows for a meaningful integration of the studies; (2) their distinct origins, which might indicate differences in mechanisms of action; (3)

chemical modifiability. The fact that those fibres are commonly used, and offer the possibility of chemically modifying them, makes it possible to consider how to combine the properties to create an optimally performing fibre.

To present the findings, we created a graph that shows the found levels of increase or decrease in LDL, HDL, VLDL and total cholesterol per study in one graph for each fibre. It also allowed us to calculate a weighted mean. A formal forest plot was not possible, as the standard deviations were usually not reported.

As discussed in 2.2.7, publication bias is a common threat to the validity of systematic research. One way of estimating if this is the cause of some of the effect that is found is a funnel plot. In this review, we found likely publication bias with a funnel plot. This means that the final finding is likely an overestimate of the real effect, because small studies with random high findings were published, but those with random low findings were not.

6.2 Full text

A preliminary version of this chapter was published in Food Chemistry.

van der Gronde, T., Hartog, A., van Hees, C., Pellikaan, H., & Pieters, T. (2016). Systematic review of the mechanisms and evidence behind the hypocholesterolaemic effects of HPMC, pectin and chitosan in animal trials. *Food Chemistry*, 199, 746–759. https://doi.org/10.1016/j.foodchem.2015.12.050

Systematic review on the mechanisms and evidence behind the hypocholesterolaemic effect of HPMC, pectin and chitosan in animal trials

Toon van der Gronde[1], Anita Hartog[1,2], Charlotte van Hees[1], Hubert Pellikaan[3], Toine Pieters[1,4]*

[1]Department of Pharmaceutical Sciences, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Utrecht, The Netherlands

[2]Nutricia Research, Uppsalalaan 12, 3583 CT Utrecht, The Netherlands

[3]Nextstep Pharma, Utrecht, The Netherlands

[4]Freudenthal Institute, Utrecht University, Utrecht, The Netherlands

*Corresponding author: t.pieters@uu.nl, tel: 0031613621033

Highlights

- Hydroxypropyl methylcellulose (HPMC), pectin and chitosan reduce total cholesterol
- Different mechanisms of these fibres give this result
- Designing a fibre with a focus on biophysical mechanisms of action is promising

6.2.1 Abstract

Dietary fibres have diverse mechanisms in reducing plasma cholesterol, which could be useful for treating high levels of low-density lipoprotein cholesterol (LDL-C). The objective of this review is to determine the state of the evidence for the cholesterol-lowering effects of three selected fibres and their mechanisms, using the most recent animal trials. Therefore, a systematic review was conducted for hydroxypropyl methylcellulose (HPMC), pectin and chitosan in Pubmed, Embase and the Cochrane Library. All fibres reviewed reduced total cholesterol, very low-density lipoprotein cholesterol (VLDL-C) and LDL-C. Pectin gave a small, and chitosan an impressive rise in high-density lipoprotein cholesterol (HDL-C). A limitation of this study is the variety of animal models, each with distinct cholesterol profiles. Possible publication bias was also detected. In conclusion, chitosan seems to be the most promising of the studied fibres. A dietary fibre could be designed that

yields the best cholesterol-lowering effect, using experiences in tailoring physicochemical properties and exploiting the primarily biophysical mechanisms of action.

6.2.2 Introduction

6.2.2.1 Relevance

In middle- and high-income countries, cardiovascular diseases are the leading causes of death and disability (Aller et al., 2004; Erkkila & Lichtenstein, 2006; Theuwissen & Mensink, 2008; Parolini et al., 2013). Relevant risk factors for these diseases are smoking and obesity (Erkkila & Lichtenstein, 2006; Choi et al., 2012), as well as a high serum concentration of low-density lipoprotein cholesterol (LDL-C) (Carr, Wood, Hassel, Bahl, & Gallaher, 2003; Aller et al., 2004; Erkkila & Lichtenstein, 2006; Maki et al., 2009b; Yokoyama et al., 2011; Marounek, Volek, Duskova, Tuma, & Taubner, 2013). The diet-heart hypothesis links an intake of high quantities of saturated fat to a high LDL-C level, which increases the risk of cardiovascular diseases (Cos et al., 2001; Erkkila & Lichtenstein, 2006). In fact, a 1% reduction in total cholesterol is associated with an approximately 2.5% reduction in coronary heart disease incidence (Choi et al., 2012). Decreasing LDL-C levels would benefit the populations in many countries.

The hypocholesterolaemic effects of specific fibres have been known for over 40 years (Aller et al., 2004; Hong, Turowski, Lin, & Yokoyama, 2007; Theuwissen & Mensink, 2008; Bartley et al., 2010; Cox et al., 2013). The literature suggests that a 10g per day increase in intake of dietary fibre can lower the risk of coronary events by 12% and deaths by 19% (Pereira et al., 2004). Most people, however, do not consume the recommended amount of fibre on a daily basis (Grizard, Dalle, & Barthomeuf, 2001; Aprikian et al., 2003; Theuwissen & Mensink, 2008; Bazzano, 2008; Anderson et al., 2009; Gallaher & Gallaher, 2009) which is considered to be 25 to 38 grams per day (Department of Agriculture., National Agricultural Library., National Academy of Sciences., Institute of Medicine., & Food and Nutrition Board., 2005; EFSA Panel on Dietetic Products, 2010a).

6.2.2.2 Cholesterol

Cholesterol in the human body originates from two sources: it is synthesized in the liver (accounting for 700-900mg per day), and taken in by diet (300-500mg per day) (Gunness & Gidley, 2010).

Small amounts of cholesterol are used for the synthesis of steroid hormones (Gunness & Gidley, 2010) and cell membranes (Park, Choi, & Kim, 2000; Gunness & Gidley, 2010). Most cholesterol is used by the liver to synthesize bile acids, such as cholic acid (Gunness & Gidley, 2010; Choi et al., 2012), which are needed to increase absorption of hydrophobic nutrients (Gunness & Gidley, 2010). Bile acids are stored in the gall bladder and released into the duodenum and proximal jejunum after being stimulated by cholecystokinin (Gunness & Gidley, 2010).

Approximately 95% of excreted bile acids are reabsorbed and recycled by reuptake (Gunness & Gidley, 2010; Choi et al., 2012). The lost fraction is compensated by synthesis and diet uptake (Gunness & Gidley, 2010).

Cholesterol is transported by lipoproteins, which are small spheres containing phospholipids, apolipoproteins and cholesterol. These lipoproteins can be divided into subgroups: LDL-C, high-density lipoprotein cholesterol (HDL-C) and very low-density lipoprotein cholesterol (VLDL-C). LDL-C is associated with cardiovascular diseases, and as such is considered 'bad' cholesterol that contributes to plaque formation (Bokura & Kobayashi, 2003; Carr et al., 2003; Aller et al., 2004; Erkkila & Lichtenstein, 2006; Maki et al., 2009b; Yokoyama et al., 2011; Marounek et al., 2013; Nordestgaard & Varbo, 2014).

Conversely, HDL-C is considered 'good' cholesterol that protects against cardiovascular diseases (Yao & Chiang, 2002; Bokura & Kobayashi, 2003; Gunness & Gidley, 2010; Ban, Rico, Um, & Kang, 2012;Rader & Hovingh, 2014), and helps scavenge LDL-C from the arteries and carry it back to the liver, where it is broken down (Hossain et al., 2007; Zong et al., 2012; Ban et al., 2012)

6.2.2.3 Cholesterol reduction

Statins, in combination with lifestyle modifications, are used as first-line drug treatment for hypercholesterolaemia (Maki et al., 2009b; Metzger, Barnes, & Reed, 2009; Parolini et al., 2013). Statins inhibit 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (HMG-CoA reductase), the rate-limiting enzyme in cholesterol biosynthesis (Maki et al., 2009b; Gunness & Gidley, 2010; Marounek et al., 2013). However, statins are often ineffective for achieving the treatment goal for LDL-cholesterol (Jaffer & Sampalis, 2007; Maki et al., 2009b; Marounek et al., 2013) . This is partially due to side effects that many patients suffer including headache,

nausea, vomiting and muscle pain (Bokura & Kobayashi, 2003; Jaffer & Sampalis, 2007; Marounek et al., 2013), causing reluctance of clinicians to use high doses (Maki et al., 2009b). Given that specific dietary fibres show cholesterol-lowering effects, they have the potential to assist in reaching the desired cholesterol-lowering target with fewer side effects (Jaffer & Sampalis, 2007; Maki et al., 2009b).

6.2.2.4 Fibres

Dietary fibres consists of non-digestible carbohydrates and lignin, including non-starch polysaccharides, that are intrinsic and intact in plants (Department of Agriculture. et al., 2005; EFSA Panel on Dietetic Products, 2010a). Primary sources of fibres include fruits, vegetables, legumes, nuts and cereal products (Erkkila & Lichtenstein, 2006; Bazzano, 2008; Anderson et al., 2009). These substances resist digestion by enzymes produced in the human gastrointestinal system (Brown, Rosner, Willett, & Sacks, 1999; Dharmarajan, Ravunniarath, & Pitchumoni, 2003; Aller et al., 2004; Bazzano, 2008; Chang, Yao, & Chiang, 2012). Whether fibres from sources other than plants can be considered dietary fibres is questionable (Dakhara, Anajwala, & Selote, 2012). These are sometimes referred to as functional fibres (Department of Agriculture. et al., 2005) or seen as dietary fibre too, given their comparable digestibility (EFSA Panel on Dietetic Products, 2010a).

Fibres can be categorised as soluble and insoluble fibres (Brown et al., 1999; Erkkila & Lichtenstein, 2006; Theuwissen & Mensink, 2008; Anderson et al., 2009). Examples of soluble fibres include gums, pectins, hydroxypropyl methylcellulose and mucilages (Aller et al., 2004; Erkkila & Lichtenstein, 2006; Bazzano, 2008; Theuwissen & Mensink, 2008; Anderson et al., 2009; Brockman, Chen, & Gallaher, 2012). Most are fermented in the large intestine (Anderson et al., 2009). Some have gel-forming properties with increasing viscosity and these are often referred to as viscous fibres (Erkkila & Lichtenstein, 2006).

Examples of insoluble fibres, or structural or matrix fibres, are cellulose and lignin (Aller et al., 2004; Erkkila & Lichtenstein, 2006; Bazzano, 2008; Theuwissen & Mensink, 2008; Bartley et al., 2010). They are fermented to a limited extent, but are beneficial because of a bulking action that promotes bowel regularity (Dharmarajan et al., 2003; Theuwissen & Mensink, 2008; Anderson et al., 2009).

The health-promoting effects of soluble and insoluble fibres differ similarly with respect to their characteristics. Insoluble fibres have been proven to reduce cardiovascular incidence more than soluble fibres (Erkkila & Lichtenstein, 2006), while soluble fibres have shown the best results in reducing cardiovascular risk factors like cholesterol levels (Brown et al., 1999; Aller et al., 2004; Erkkila & Lichtenstein, 2006; Eussen et al., 2010; Parolini et al., 2013). This discrepancy is not yet understood.

6.2.2.5 Mechanisms of action

Literature describes various proposed mechanisms of action by which fibres are thought to influence plasma cholesterol values. A brief overview of these mechanisms is show in figure 1.

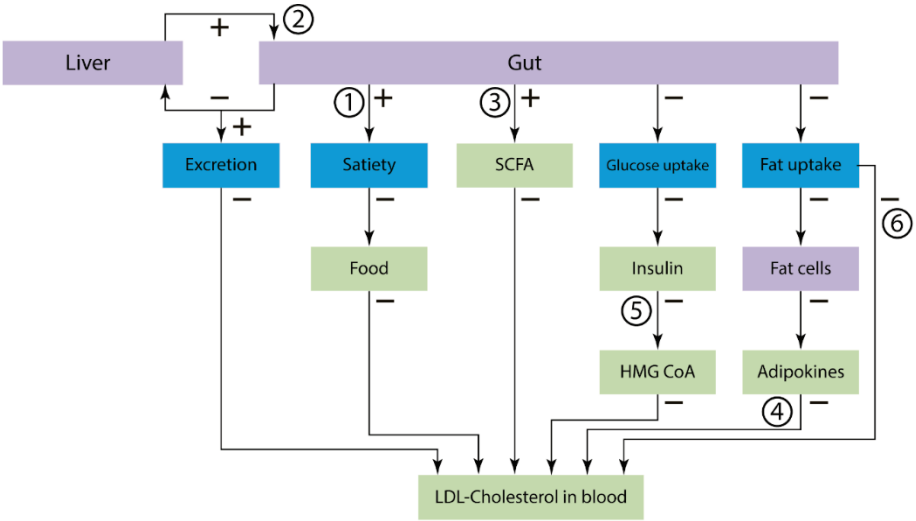

Figure 1. Schematic overview of the proposed mechanisms of the hypocholesterolaemic effects of fibres. The influence of fibres on each process is indicated with a ''+'' (increasing due to fibre intake) or ''-'' (decreasing due to fibre intake) sign. A more detailed description of each mechanism is given in section 6.2.2.5 Mechanisms of action, to which the numbers in the figure refer.

The first proposed mechanism points to the fact that fibres are able to decrease overall energy intake (Brown et al., 1999; van Bennekum, Nguyen, Schulthess, Hauser, & Phillips, 2005; Eussen et al., 2010). Food high in fibres often contains fewer calories than the nutrient-dense food it replaces, and takes longer to ingest and digest than its low-fibre counterparts (Erkkila & Lichtenstein, 2006; Anderson et al., 2009).

Soluble fibres increase viscosity and bind water, which increases bulk forming, and thus satiety. Fibres also delay gastric emptying and slow down food's progress through the small intestine, which promotes satiety (van Bennekum et al., 2005; Anderson et al., 2009; Gunness & Gidley, 2010; Eussen et al., 2010; Krzysik, Grajeta, Prescha, & Weber, 2011; Brockman et al., 2012). In the small intestine, dietary fibres modulate incretin secretion, which stimulates insulin release (Anderson et al., 2009) and can also reduce appetite (Anderson et al., 2009; Brockman et al., 2012). In sum, increased satiety and reduced appetite can reduce consumption of cholesterol-containing food, leading to lower cholesterol levels.

The second mechanism of action is an inhibition of bile acids reuptake (Anderson et al., 2009; Bartley et al., 2010; Hur, Lee, & Lee, 2015). Bile acids are synthesised in the liver from cholesterol (Bartley et al., 2010). They facilitate the absorption of cholesterol, lipids and vitamins that are soluble in intestinal fat (Bartley et al., 2010). By entrapping bile acids through means of viscosity (Grizard et al., 2001; Gunness & Gidley, 2010) or adsorption (Theuwissen & Mensink, 2008; Gunness & Gidley, 2010; Parolini et al., 2013) and inhibiting micelle formation (Grizard et al., 2001; Marounek, Volek, Synytsya, & Copikova, 2007; Gunness & Gidley, 2010; Brouns et al., 2012; Hur et al., 2015), soluble fibres increase faecal excretion of cholesterol and bile acids (Erkkila & Lichtenstein, 2006; Anderson et al., 2009; Eussen et al., 2010; Hung, Anderson, Albers, Langhorst, & Young, 2011; Hur et al., 2015). Other research suggests that soluble fibres form a viscous layer of water in the lumen, thus preventing the permeation of bile acids and cholesterol (Erkkila & Lichtenstein, 2006; Theuwissen & Mensink, 2008; Maki et al., 2009c; Eussen et al., 2010). Both mechanisms lead to an inhibited reuptake of bile acids in the small intestine, decreasing the enterohepatic pool of bile acids (Dongowski & Lorenz, 2004; Erkkila & Lichtenstein, 2006; Brufau, Canela, & Rafecas, 2007; Sanchez et al., 2008; Bartley et al., 2010). This decrease needs to be compensated; therefore, hepatic enzymes such as cholesterol 7-α-hydroxylase, also known as CYP7A1 (the rate limiting enzyme involved in bile acid production (Zong et al., 2012) are upregulated (Marounek et al, 2007; Bartley et al., 2010; Eussen et al., 2010; Kim et al., 2011; Zong et al., 2012). CYP7A1 uses cholesterol, leading to lower cholesterol levels in liver cells (Ausar et al., 2003; Eussen et al., 2010). This, in turn, upregulates LDL-receptors (Bazzano, 2008; Zong et al., 2012) and CYP51 and HMG-CoA reductase

production in the liver cells (Eussen et al., 2010; Bartley et al., 2010), which increases uptake of LDL-cholesterol from the blood (Erkkila & Lichtenstein, 2006; Theuwissen & Mensink, 2008; Eussen et al., 2010).

This second mechanism can also be stimulated in a more direct way. Some fibres can increase bile acid production from cholesterol by upregulating hepatic LDL receptor mRNA, enhancing LDL uptake from the blood (Anandan et al., 2013).

The third mechanism is the inhibition of cholesterol synthesis by short-chain fatty acids (SCFAs), such as propionate, acetate and butyrate, which are products of soluble fibre fermentation from intestinal bacteria (Bazzano, 2008; Theuwissen & Mensink, 2008; Bartley et al., 2010; Eussen et al., 2010; Gunness & Gidley, 2010; Krzysik et al., 2011; Brockman et al., 2012; Santas, Espadaler, Mancebo, & Rafecas, 2012). After absorption, each SCFA has a specific function: acetate can be used as a substrate for cholesterol synthesis (Yao, Huang, & Chiang, 2008), propionate inhibits this process through HMG-CoA reductase inhibition (Dvir et al., 2000; Maki et al., 2000; Park et al., 2000; Terpstra, Lapre, De Vries, & Beynen, 2002b; Dongowski & Lorenz, 2004; Brufau et al., 2007; Bazzano, 2008; Bartley et al., 2010; Gunness & Gidley, 2010), and butyrate serves as an energy source for enterocytes (Park et al., 2000; Marounek et al., 2007). However, evidence suggests that HMG-CoA reductase inhibition is not the primary mechanism (Gunness & Gidley, 2010), and studies in humans are scarce and contradictory (Theuwissen & Mensink, 2008; Gunness & Gidley, 2010).

SCFAs also stimulate intestinal hormone secretions, for example, peptide tyrosine-tyrosine (PYY) and glucagon-like peptide-1 (GLP-1), which increase satiety (Brockman et al., 2012). Furthermore, SCFAs lower colonic pH, which decreases bile acid solubility, thus preventing reabsorption (Dongowski & Lorenz, 2004).

The fourth proposed mechanism is found in adipokines, a class of substances produced by fat cells. High fat diets result in an increase of leptin, resistin and TNF-α levels and a decrease of adiponectin (Hung et al., 2009; Ban et al., 2012; Liu et al., 2012). Adipokines are important for glucose regulation and lipid metabolism, including cholesterol (Hung et al., 2009; Bartley et al., 2010; Ban et al., 2012; Brockman et al., 2012; Hsieh, Yao, Cheng, & Chiang, 2012).

A more beneficial profile of these adipokines and cholesterol has been found in hamsters and mice with high fibre intake as compared to lower fibre intake (Hung et al., 2009; Ban et al., 2012). The proposed mechanism is the reduction of adipose tissue, leading to a more advantageous adipokine production and secretion (Ban et al., 2012), which in turn is thought to improve plasma cholesterol concentration.

The fifth mechanism indicates that fibres delay intestinal absorption of glucose, due to increased viscosity of the intestinal content, which leads to lower insulin levels (Schwab, Louheranta, Torronen, & Uusitupa, 2006; Sanchez et al., 2008; Theuwissen & Mensink, 2008; Gunness & Gidley, 2010). Insulin activates HMG-CoA reductase, therefore lower insulin could result in lower cholesterol production (Theuwissen & Mensink, 2008; Gunness & Gidley, 2010). Whether this effect is relevant has not been proven, but increases in levels of HMG-CoA reductase have been shown in human trials (Theuwissen & Mensink, 2008).

The sixth mechanism relies on fatty acids. Lauric, myristic and palmitic acid (C12:0, C14:0, C16:0 respectively) decrease LDL-receptor production and activity, which leads to an increase in LDL-C (Fernandez & West, 2005; Santas et al., 2012). Inhibition of the absorption of these fatty acids through fibre intake would lower LDL-C (Santas et al., 2012).

Likely, multiple mechanisms are valid for each fibre, but the relevance of the proposed mechanisms varies. Current research is focussed on determining the importance of each mechanism for all dietary fibres (Carr et al., 2003; Hung et al., 2009; Eussen et al., 2010;).

To determine the state of the evidence of the cholesterol-lowering effects of fibres and their mechanisms, three specific fibres will be reviewed in this article: the semisynthetic hydroxypropyl methylcellulose (HPMC), pectin (originates from fruits and vegetables), and chitosan (found in shellfish).

6.2.3 Method

HPMC, chitosan and pectin were selected for review based on 1) the evidence of their cholesterol-lowering effect. Many studies on these fibres have been published, so information is likely to be available, 2) their distinct origins, which might indicate differences in mechanisms of action and 3) chemical modifiability.

While meta-analyses in human populations have already been conducted, they don't provide information on the mechanisms behind the observed effect. However, animal research provides for more measurements and can provide more insight into the relationship between the structure of a fibre and its effect. Therefore, this study performed a quantitative analysis on animal studies only.

6.2.3.1 Objective

The aim of this review is to gain insight in the mechanisms that cause the cholesterol-lowering effect of selected fibres and present an overview of the effects that are seen in different animal models. This could also provide recommendations for future research.

6.2.3.2 Data sources

A systematic review was conducted using articles retrieved from Pubmed, Embase and the Cochrane Library. ''[Fibre name] AND cholesterol'' was used as the search string, on September 2, 2014. Publications were included in the systematic search if they were published in peer-reviewed journals between January 1, 2000 and the date of the search. No further limits were used.

6.2.3.3 Exclusion criteria

- In vitro studies
- Studies done in non-vertebrate animals
- Studies in which the fibres were not individually evaluated
- Articles published in languages other than English
- Articles published before January 1, 2000
- All types other than primary peer-reviewed articles

6.2.3.4 Study selection

Articles were assessed for their eligibility by reviewing the title and abstract. When further information was needed to determine relevance, the full text was evaluated. Additional articles were selected for inclusion in the review where deemed necessary, for example to provide information on the chemical structure of the fibres or to clarify the relevance of cholesterol. Figure 2 provides an overview of the selection procedure.

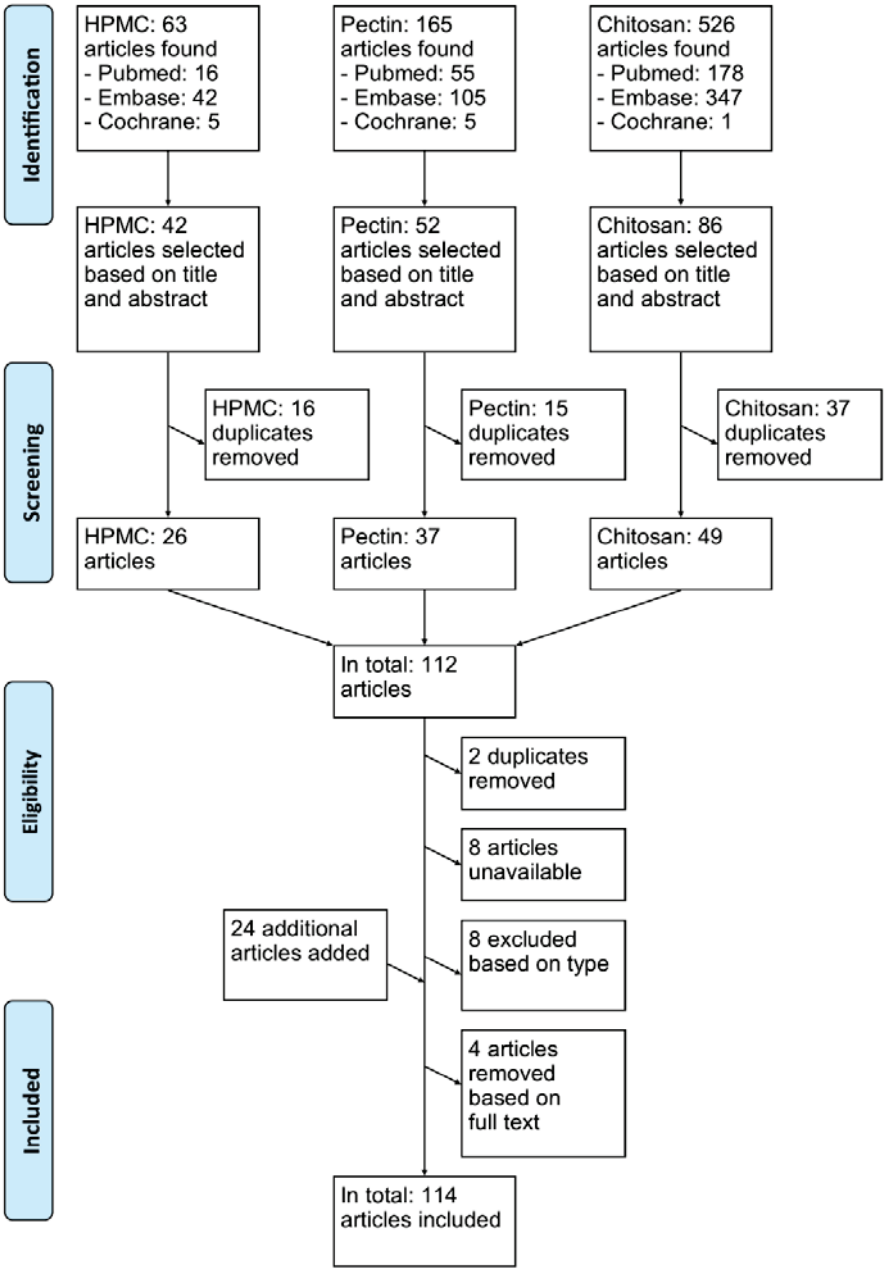

Figure 2. Schematic overview of the study selection process.

6.2.3.5 Study analysis

To analyse the results of each study, the relative difference in total cholesterol, LDL-C, HDL-C and VLDL-C levels between the group on a high-fat diet and the group fed with a high-fat diet enriched with one specified fibre was calculated, where data was available. The weighted mean was calculated by correcting for the number of animals that had been included in each study. These results were standardized for the relative quantity of fibre added to the diet, so that the figures give a consistent estimate of the effect a 1% fibre enriched diet would have had. To assess the possibility of publication bias, a funnel plot based on outcomes versus sample size was drawn. Microsoft Excel was used for all calculations and graphs.

6.2.4 Results

6.2.4.1 HPMC

Hydroxypropyl methylcellulose (HPMC) is a non-fermentable semi-synthetic dietary fibre based on cellulose (Burdock, 2007), which is a carbohydrate based on anhydroglucose units (Reppas, Swidan, Tobey, Turowski, & Dressman, 2009). It is commonly used as a stabiliser, emulsifier and thickener (Burdock, 2007; Hung et al., 2011; Ban et al., 2012). After ingestion, HPMC forms a viscous solution in the gastrointestinal tract (Maki et al., 2000; Burdock, 2007; Maki et al., 2009a; Maki et al., 2009c).

6.2.4.1.1 Mechanisms

HPMC is non-fermentable due to its cellulose backbone (Hung et al., 2009); it does not yield SCFA production. In animal studies, a relationship has been found between the viscosity of used HPMC and a hypocholesterolaemic effect of bile acid and cholesterol excretion (Bartley et al., 2010). Therefore, HPMC's action is thought to increase the viscosity of intestinal content (Hung et al., 2009; Bartley et al., 2010; Ban et al., 2012), which increases faecal excretion of bile acids (Hung et al., 2011;Yokoyama et al., 2011; Cox et al., 2013) (second mechanism).

This action can cause either direct adsorption of bile acids on HPMC, or the formation of a barrier that decreases permeability (Maki et al., 2000; Maki et al., 2009a; Yokoyama et al., 2011; Ban et al., 2012). This last option was shown when in vivo permeability was determined using a fluorescein isocyanate-labelled dextran polymer: the polymer was decreased by 50% in the group taking HPMC supplementation (Kim,

Bartley, Young, Seo, & Yokoyama, 2013). This also suggests that the timing of HPMC consumption is important; ingesting HPMC during a meal increases its effectiveness (Maki et al., 2000).

The mRNA level of CYP7A1 (involved in bile acid synthesis (Kim et al., 2011)) was found to be upregulated after HPMC administration (Kim et al., 2013). This is to be expected, since bile acid production needs to increase to compensate for the lower reuptake.

HPMC has been shown to have a beneficial effect in regulating adipokine production (Bartley et al., 2010; Ban et al., 2012). High fat diets resulted in an increase of leptin, resistin and TNF-α levels and a decrease of adiponectin, but the addition of HPMC counteracted this outcome (Hung et al., 2009; Ban et al., 2012). This suggests that the cholesterol-lowering effects of HPMC could also partially be due to a decrease in adipose tissue through diminished absorption of fat and lower food intake. This would lead to a more advantageous production of adipokines, thus leading to better cholesterol levels (Hung et al., 2009; Ban et al., 2012) (fourth mechanism).

6.2.4.1.2 Animal studies

Several studies in animals have shown the hypocholesterolaemic effects of HPMC, which can cause weight reduction (Kim et al., 2013). The results of the eight studies that were found and selected in this search are given in table 1 (in the supplementary material) and figure 3.

6.2.4.1.3 Human studies

In human subjects, HPMC has been tested in doses of 5-30g per day (Maki et al., 2000; Maki et al., 2007; Anderson et al., 2009; Maki et al., 2009a; Reppas et al., 2009; Cox et al., 2013). Various clinical trials have indicated that HPMC lowers LDL-C (Maki et al., 2000; Anderson et al., 2009; Maki et al., 2009a; Maki et al., 2009c; Reppas et al., 2009) without changing HDL-cholesterol (Anderson et al., 2009; Maki et al., 2009a; Reppas et al., 2009) or triglyceride concentrations (Anderson et al., 2009; Maki et al., 2009a), and has an hypoglycaemic effect (Maki et al., 2007). An 8.5% reduction in LDL-C has been found using 5g per day of HPMC (Anderson et al., 2009). The highly viscous HPMC was more effective than its low-viscous counterpart in human studies (Reppas et al., 2009).

Another trial was performed to determine whether adding HPMC to statin therapy would prove beneficial. The addition of HPMC resulted

in a total reduction in LDL-C (10.5%) and total cholesterol (7.4%), compared to statin therapy alone (Maki et al., 2009b).

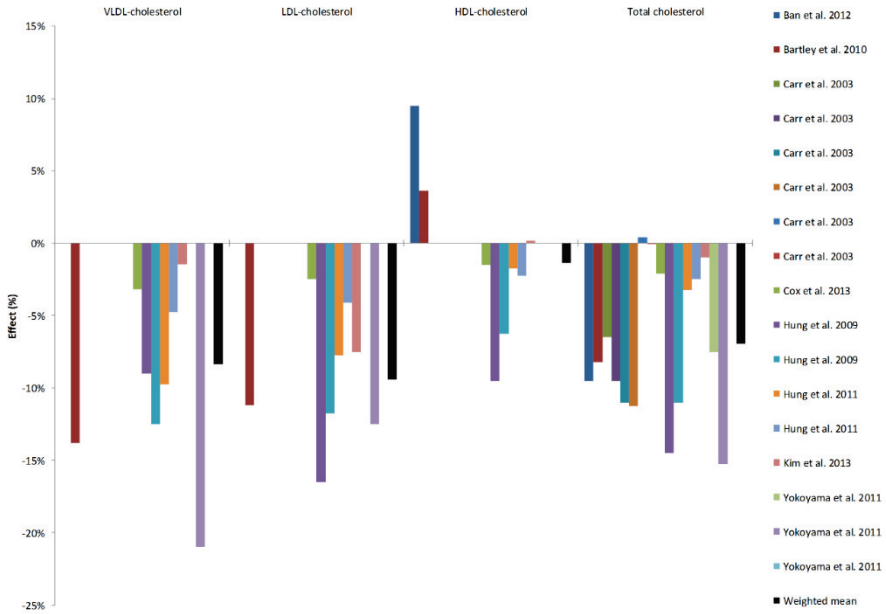

Figure 3. The average effect of HPMC on cholesterol levels by percent administered in diet. To correct for differences in the amount of fibre added to the diet, all results are standardised so that the figures correspond to 1% of the diet being fibre. All reviewed studies, including not significant results, are displayed. The weighted mean was calculated for each cholesterol class using all studies that reported on that specific class.

6.2.4.1.4 Conclusion

HPMC has only mild side effects (Maki et al., 2000) and effectively decreases total and LDL-C levels in animals and humans. It might be useful as an additive in food products for human use, or as an add-on to statin therapy, as demonstrated in a human trial. (Maki et al., 2009b). However, HPMC lacks the ability to increase HDL-C. This casts doubts on the actual health-promoting effect of HPMC, namely whether it can actually prevent cardiovascular diseases. Further research on hard outcomes, like the effect on coronary events and death, needs to be performed before conclusions can be made.

6.2.4.2 Pectin

Several subtypes of pectin exist, but all are a polymer consisting of blocks of polar D-galacturonic acid and rhamnogalacturonan (Thakur, Singh, & Handa, 1997). It is found in the plant cell walls of fruits and vegetables (Thakur et al., 1997). Pectin can be used as a fat replacement in food (Marounek et al., 2007; Theuwissen & Mensink, 2008; Brouns et al., 2012) . The molecular weight and the degree of esterification of the carboxyl groups determine its physiochemical properties (e.g. whether it dissolves (Thakur et al., 1997; Dongowski & Lorenz, 2004; Marounek et al., 2007), and its effectiveness in lowering cholesterol (Marounek et al., 2007; Brouns et al., 2012). If more than 50% of the galacturonic acid residues are esterified it is called high methoxyl pectin (Theuwissen & Mensink, 2008). Pectins with a high degree of hydrophobic substitution are less soluble (Marounek et al., 2007), and are thought to be more effective in lowering plasma cholesterol (Dongowski & Lorenz, 2004; Brufau et al., 2007; Marounek et al., 2007; Sanchez et al., 2008).

6.2.4.2.1 Mechanisms

The fact that the cholesterol-lowering effect of pectin depends on the ability to form a viscous gel (Brouns et al., 2012) is an important clue about the mechanism of action. It is assumed that pectin binds to cholesterol and bile acids in the gut, both reducing reabsorption and promoting their excretion (Park et al., 2000; Dongowski & Lorenz, 2004; Brufau et al., 2007; Marounek et al., 2007; Hur et al., 2015). Pectin also disturbs micelle formation (Dongowski & Lorenz, 2004), thus inhibiting absorption of cholesterol (Brufau et al., 2007) (second mechanism).

Pectin is not degraded by enzymes during passage through the human small intestine (Dongowski & Lorenz, 2004; Marounek et al., 2007). In the large intestine, however, it is almost completely fermented to SCFAs (Dongowski & Lorenz, 2004; Marounek et al., 2007; Kirat, 2010) of which acetate is most predominant (Thakur et al., 1997; Marounek et al., 2007; Kirat, 2010). Acetate is a cholesterol precursor (Marounek et al., 2007; Kirat, 2010), whereas propionate inhibits hepatic cholesterol synthesis (Dongowski & Lorenz, 2004). These data indicate that the pectin-induced changes in an acetate:propionate ratio are unlikely to significantly contribute to pectin's cholesterol-lowering effect (third mechanism).

It is recognised that pectin reduces the blood glucose rise after a meal (EFSA Panel on Dietetic Products, 2010b). This results in lower

insulin levels, leading to lower HMG-CoA reductase activity (Gunness & Gidley, 2010; Theuwissen & Mensink, 2008) (fifth mechanism).

6.2.4.2.2 Animal studies

Various types of pectin have been used in animal studies, with differing results (Terpstra et al., 2002b; Marounek et al., 2007; Krzysik et al., 2011). An overview of the results is given in table 2 (in the supplementary material) and figure 4.

6.2.4.2.3 Human studies

The effect of pectin ingestion by humans has been evaluated (Schwab et al., 2006; Brouns et al., 2012). A meta-analysis of the effect of pectin in humans showed a 13% reduction of LDL-C when consuming 12-24g of pectin per day (Anderson et al., 2009). The type of pectin seems crucial because highly viscous and highly esterified pectin appears to be more effective (Brouns et al., 2012).

The European Panel on Dietetic Products, Nutrition and Allergies concluded that consumption of 6g of pectin per day contributes to the maintenance of normal cholesterol levels (EFSA Panel on Dietetic Products, 2010b).

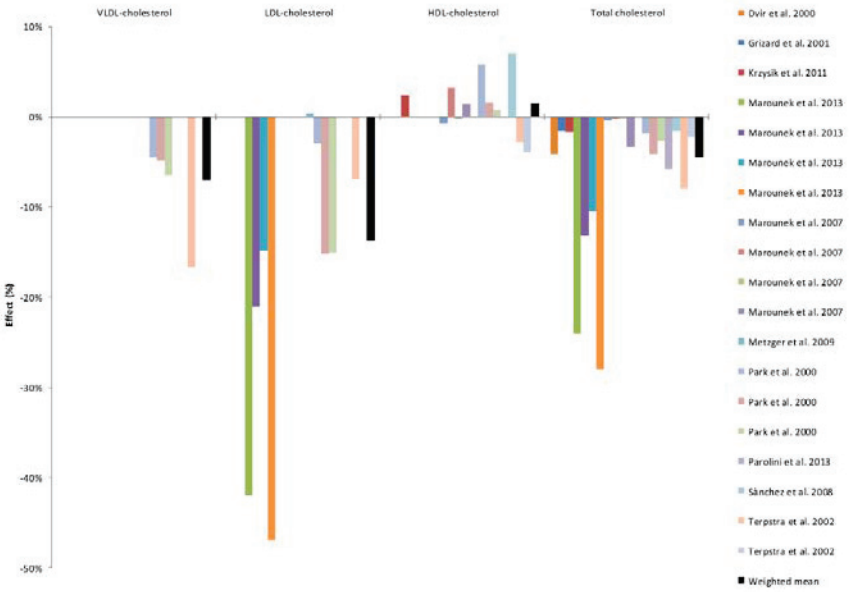

Figure 4. The average effect of pectin on cholesterol levels by percent administered in diet. All reviewed studies that mentioned the outcome,

including not significant results, are displayed. The weighted mean was calculated for each cholesterol class using all the studies that reported on that specific class.

6.2.4.2.4 Conclusion

The reviewed literature shows a diverse pattern of lowering cholesterol, which can be explained by the multiple types of pectins tested and the specific animal models. The molecular composition of pectin used for human consumption is thought to be crucial for its gel-forming properties, thus for the cholesterol-lowering effect (Terpstra, Lapre, De Vries, & Beynen, 2002a; Marounek et al., 2007; Sanchez et al., 2008; Brouns et al., 2012). Not all of the studies we included mentioned the supplement's characteristics, and our analysis did not test for this, so the cholesterol-lowering effect cannot be concluded based on our data set.

Overall, pectin seems to reduce LDL-C and total cholesterol, and possibly VLDL-C, but the effect on HDL-C remains uncertain. Although the consumption of some pectins appears to make HDL-C levels rise, this effect is small and the clinical significance is unclear. Again, more research is needed before the health effect of pectins on humans can be concluded.

6.2.4.3 Chitosan

Second to only cellulose, chitosan is the most abundant and easily obtained natural polymer known (Hossain et al., 2007; Rizzo et al., 2013). It is found in shellfish like clams and krill oysters, squid, fungi, yeasts and insects (Baker, Tercius, Anglade, White, & Coleman, 2009; Hsieh et al., 2012; Anandan et al., 2013). Soluble chitosan is derived from insoluble chitin by de-N-acetylation and a polymer of N-acetyl-D-glucosamine and D-glucosamine (Tharanathan & Kittur, 2003). Relevant parameters for chitosan's ability to lower cholesterol are molecular weight (Chang et al., 2012), degree of acetylation – deacetylation makes chitosan less susceptible to degradation (Yao & Chiang, 2006a; Yao et al., 2008) – and pKa – which determines the ability to dissolve (Ylitalo, Lehtinen, Wuolijoki, Ylitalo, & Lehtimaki, 2002; Lee, Park, Choi, Yi, & Shin, 2003; Ni et al., 2004; Yao & Chiang, 2006a; Yao et al., 2008; Hernandez-Gonzalez, Gonzalez-Ortiz, Martinez-Abundis, & Robles-Cervantes, 2010; Choi et al., 2012; Liu et al., 2012). Viscosity appears to be less relevant (Chiang, Yao, & Chen, 2000).

Chitosan is strictly not a dietary fibre, because it does not have a vegetable origin, but it does have the same chemical and physiological properties as vegetable fibres (Tharanathan & Kittur, 2003; Sumiyoshi & Kimura, 2006; Yao & Chiang, 2006a; Zhang, Liu, Li, & Xia, 2008). Since chitosanase is not produced in most animal intestines (Kim, Park, Yang, & Han, 2001), chitosan is indigestible for mammals (Gallaher et al., 2000; Ausar et al., 2003; Asha & Nair, 2005; Sumiyoshi & Kimura, 2006). However, chitosan oligosaccharide, the soluble hydrolysis product of chitosan, is water-soluble and can be partially digested and absorbed by mammals (Kim et al., 2001; Wang et al., 2011; Zong et al., 2012).

6.2.4.3.1 Mechanisms

Chitosan is thought to act in several ways. First, where chitosan is soluble, it lowers cholesterol levels by increasing the viscosity of stomach content (Gallaher et al., 2000; Ausar et al., 2003), which inhibits uptake of cholesterol (second mechanism). This action delays gastric emptying (Chang et al., 2012), thus leading to a decrease in food intake by inducing satiety (van Bennekum et al., 2005) (first mechanism).

Next, chitosan acts as a cationic polysaccharide in an acidic environment like the stomach (Ylitalo et al., 2002; Tang et al., 2005), so the positive amino groups of the fibre bind to negatively charged molecules, such as bile acids and fatty acids (Tai, Sheu, Lee, Yao, & Chiang, 2000; Kim et al., 2001; Ausar et al., 2003; Jaffer & Sampalis, 2007; Baker et al., 2009; Anraku et al., 2010; Choi et al., 2012; Jun et al., 2012; Anandan et al., 2013). This leads to higher activity of the LDL-receptor and thus lower LDL-C plasma levels (Santas et al., 2012) (sixth mechanism). Neutrally charged triglycerides are not affected here (Hossain et al., 2007).

In the intestine, a higher pH makes the complex precipitate with bound fatty and bile acids and cholesterol (Tang et al., 2005; Yao & Chiang, 2006b; Zhang et al., 2008; Zhang, Zhang, Mamadouba, & Xia, 2012). After precipitation, the bound fatty- and bile acids are inaccessible to enzymes (Tai et al., 2000; Hossain et al., 2007) and are excreted with the stool (Ausar et al., 2003; Sumiyoshi & Kimura, 2006; Baker et al., 2009; Anraku et al., 2010; Choi et al., 2012; Anandan et al., 2013). The lack of cholesterol impairs emulsifications and decreases triglyceride uptake (Ylitalo et al., 2002; Jun et al., 2012). In vitro, chitosan can bind approximately four times its own weight in lipids (Guha et al., 2005).

Another way of action could be dependent on chitosan's antioxidant activity. Removal of abnormalities in lipid metabolism that are associated with oxidative phenomena creates better cholesterol values (Anraku et al., 2009; Anraku et al., 2010). Reduced oxidative stress has been found due to chitosan in animals (Anraku et al., 2010; Ahmed et al., 2014) and humans (Anraku et al., 2009). However, other studies indicate that chitosan causes increased oxidative stress (Hossain et al., 2007) advising antioxidant supplementation. This requires more research.

Finally, the inhibition of pancreatic lipase activity has been shown in vitro and in mice (Sumiyoshi & Kimura, 2006; Choi et al., 2012). This decreases lipid and cholesterol absorption. Chitosan's inhibition of liver and plasma lipase has also been proposed as a relevant mechanism (Zhang et al., 2008).

6.2.4.3.2 Animal studies

In animal studies, chitosan has been shown to be effective in reducing LDL-C and total plasma cholesterol and increasing HDL-C, and its safety has been proven (Zhang, Zhong, Tao, Wu, & Su, 2012). An overview of the results is presented in table 3 (in the supplementary material) and figure 5.

6.2.4.3.3 Human studies

Multiple studies have tested the effect of chitosan on cholesterol in humans. Some have found positive effects (Ylitalo et al., 2002; Bokura & Kobayashi, 2003; Ni et al., 2004; Jaffer & Sampalis, 2007; Choi et al., 2012), but others none (Hernandez-Gonzalez et al., 2010; Ho, Tai, Eng, Tan, & Fok, 2001). A meta-analysis conducted solely in patients with hypercholesterolaemia – those who would benefit from the effect most – found a significant reduction of total cholesterol, although LDL-C and HDL-C were not significantly affected (Baker et al., 2009). A Cochrane review included various studies and found that chitosan produced a significant increase in HDL-C and a reduction of LDL-C and total cholesterol (Jull, Ni, Bennett, Dunshea-Mooij, & Rodgers, 2008); however, clinical relevance was questioned.

To test whether chitosan actually binds fat in the human intestine, as it does in animal models, several studies measured faecal fat excretion. Results varied as some found an effect while others did not (Gallaher et al., 2002; Ni et al., 2004). Bile acid excretion, however, was increased in all studies (Gallaher et al., 2002).

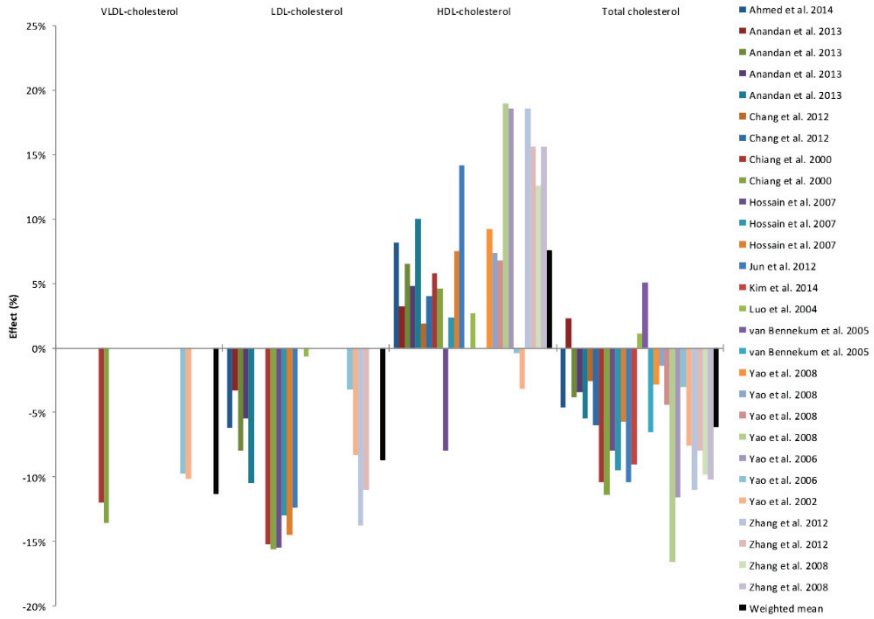

Figure 5. The average effect of chitosan on cholesterol levels by percent administered in diet. All reviewed studies, including not significant results, are displayed. Studies that administered through water are not displayed. The weighted mean was calculated for

6.2.4.3.4 Conclusion

The use of many types of chitosan with various properties complicates the comparison of studies. However, the claim that chitosan has a positive influence on the cholesterol profile in animals and humans is supported by the presented data. The high HDL-C increase is particularly impressive. The downside of high chitosan doses is the diminished uptake of essential fatty acids, fat-soluble vitamins and minerals (Kim et al., 2001). It is hypothesised that chitosan oligosaccharide would not have this unwanted effect (Kim et al., 2001), but there are not enough data available to support this hypothesis.

6.2.4.4 Other fibres

As previously stated, this paper provides a systematic but time-limited overview of the field of fibre research. Far more studies than the ones mentioned in this article have been performed. Other fibres, like psyllium, polysaccharides like inulin, polydextrose, β-glucan, and guar gum, as well as fibre rich foods such as corn bran and oat bran are not

discussed in this article, although they too might have beneficial effects for lowering cholesterol (Choi et al., 2012; Dakhara et al., 2012).

Most of those studies reviewed confirm the cholesterol-lowering effects in humans of the fibres under survey in this review. In general, 6g of water soluble viscous fibre daily intake was found to lower LDL-C by approximately 5.4% (Anderson et al., 2009). An analysis on hard outcome parameters for fibre use such as coronary events (a 14% lower risk after an increase of 10g of dietary fibres per day) and mortality (a 27% lower risk) has even been performed (Pereira et al., 2004). These results were independent of other dietary factors, age, sex, baseline BMI, smoking, blood pressure, diabetes and hypercholesterolaemia.

6.2.5 Discussion

Pereira et al. (2004) calculated, using a pooled analysis of cohort studies, that for each 10gr per day increment in total, cereal, or fruit fibre intake, coronary risk decreased by 10% to 30%. The struggle in these calculations is correcting the results for bias; humans with high fibre intake are more likely to lead a healthier life, with more physical activity, less smoking, less fat and more vitamin intake, which yields confounding factors (Erkkila & Lichtenstein, 2006). Testing specific fibres in real-life situations requires subjects to use the dietary fibres for longer periods than currently investigated, as studies have shown that the cholesterol-lowering effect builds up over time (Anderson et al., 2009).

6.2.5.1 Limitations

Animal models for studying the cholesterol-lowering effect of fibres and the mechanisms thereof offer several benefits over human studies. In animal models, environment and food intake are controllable and costs are relatively low. However, extrapolation of the experimental outcomes to humans requires careful translational interpretations. In rodents, the majority of cholesterol is carried in HDL rather than LDL, which is the case for humans (Vitic & Stevanovic, 1993; Bergen & Mersmann, 2005; Xiangdong et al., 2011). Rodents show relatively limited plasma cholesterol changes as a reaction to fibres, and they are more resistant to atherosclerosis than humans (Jokinen, Clarkson, & Prichard, 1985). As a replacement parameter for plasma cholesterol in rats, liver cholesterol can be analysed (Gallaher et al., 2000; Aprikian et al., 2003; Carr et al., 2003; Marounek et al., 2007). In addition to rats and mice, hamsters, guinea pigs and swine are used; the advantage of these animals is that their lipoprotein metabolism is more comparable to humans (Spady &

Dietschy, 1983; Fernandez, Wilson, Conde, Vergara-Jimenez, & Nicolosi, 1999).

The use of various animal species further complicates comparison, but also enables the possibility of studying the effects of fibres on multiple mechanisms in more detail (if the same fibre is used in both animal species). The overview of the reviewed HPMC articles (table 1) confirms a difference in cholesterol metabolism between rats/mice and hamsters. It shows that the effect size in rodent serum is relatively low (0-20% decrease in total cholesterol) compared to the studies in hamsters (24-61% decrease in total cholesterol). The pectin (table 2) and chitosan (table 3) studies are mainly performed in rodents, so no comparison with hamsters can be made. In studies with hamsters, the use of males is preferred over females because of their greater responsiveness (Robins et al., 1995). In humans this difference has not been found (Pereira et al., 2004).

LDL can be divided into seven subclasses, based on size, density, physicochemical composition, metabolic behaviour and atherogenicity. Over one hundred studies have found that specifically small, dense LDL (sdLDL) is associated with cardiovascular risk (Gazi, Tsimihodimos, Tselepis, Elisaf, & Mikhailidis, 2007; Mikhailidis et al., 2011). Also, the main apolipoproteins present in LDL are apoB and apoE, while HDL has the major apolipoprotein apoA-I. The exact composition of apolipoproteins influences the class and metabolism of cholesterol (Wang et al., 2011). Although many studies have measured the various types of cholesterol (HDL-C, LDL-C, VLDL-C), subclasses are not often measured. It has been shown that not all types of LDL have the same predictive power for cardiovascular risk assessment (Gazi et al., 2007; Mikhailidis et al., 2011; Rizzo et al., 2013), so using an undivided LDL-C measurement will not give the most accurate prediction for effects on mortality or morbidity endpoints. By measuring the specific cholesterol subclasses, the predictive power of the surrogate parameter for cholesterol could be improved. There are indications that similar subclasses could be measured in animals, but further research is needed to establish a firm base for translational studies. This review does not take the various types of LDL into account, given that very few studies report on them. There are indications, however, that dietary fibres influence the composition of LDL particles (Fernandez, Abdel-Fattah, & McNamara, 1993; Fernandez et al., 1999).

6.2.5.2 Bias

The present article points out the benefits of published studies in animals, but cannot exclude possible publication bias. An indication of the likelihood of this phenomenon can be found by creating a funnel plot, plotting the outcome of each study against the sample size. Studies with fewer subjects should lead to variable results, while larger studies should have more consistent results, thus creating the shape of a funnel (Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006). The funnel plot of the selected trials is shown in figure 6. Given that the studies with larger sample sizes are probably closest to the real effect, and smaller studies should spread around that effect based on coincidence, in the obtained figure a gap is found where the small studies should report negative outcomes. For all three fibres, this means that the cholesterol-lowering effect might be slightly overestimated.

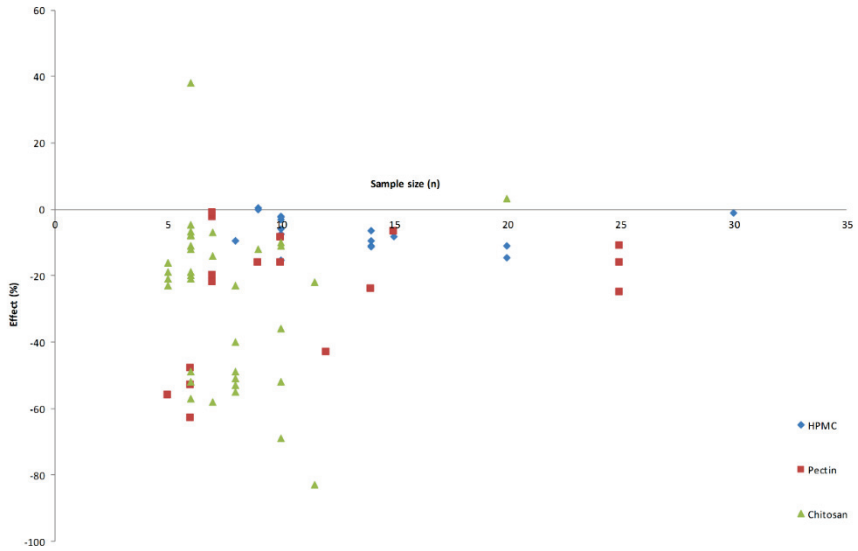

Figure 6. Funnel plot for the reviewed studies: the number of subjects per arm plotted against the found effect on total cholesterol relative to the high-fat group. For LDL-C, HDL-C and VLDL-C the same shape was found, but with a lower density because fewer studies measure these outcomes.

6.2.5.3 Combination with statins

Statins are believed to reduce total cholesterol by 25-40% (Baker et al., 2009). However, the reduction caused by fibres in humans is much lower. Fibres may be beneficial as an add-on to statin therapy, since they could further reduce plasma cholesterol by mechanisms different from statins (Guha et al., 2005; Maki et al., 2009b; Eussen et al., 2010). However, this would require more research into the interaction between fibres and statins. Fibre drug interactions have been shown in the case of chitosan and warfarin (Rizzo et al., 2013), pectins and gums combined with digoxin and acetaminophen (Dharmarajan et al., 2003), and wheat bran together with levothyroxine (Dharmarajan et al., 2003). With atorvastatin therapy, however, higher HDL-levels (Ho et al., 2001; Guha et al., 2005) and weight loss have been found by adding chitosan (Guha et al., 2005). Also, HPMC and several statins have been combined well in one study (Maki et al., 2009b). This shows that these specific combinations could be advantageous for treatment of hypercholesterolaemia.

Another point of attention in combining statins and fibres is that both modulate bile acid metabolism. Statins are indicated for decreasing bile acid production, while fibres might decrease bile acid reuptake. A lack of bile salts can cause side effects like decreased fat-soluble vitamin uptake. Moreover, some fibres are indicated for blunting HDL-cholesterol, thus increasing the effects of statins (Eussen et al., 2010; Erkkila & Lichtenstein, 2006).

6.2.6 Conclusion

The overview of the results of the selected studies shows that all three fibres under review are capable of reducing LDL-C, VLDL-C and total cholesterol. HPMC and pectin had limited or inconclusive results for HDL-C modulation, which is in line with the findings in human studies (Anderson et al., 2009; Reppas et al., 2009; Maki et al., 2009a). However, chitosan increased HDL-C despite the high basal HDL-C levels present in rats (the animal used in most of the studies).

An inverse association between fibre intake and coronary heart disease risk underlines the importance of adequate fibre consumption (Pereira et al., 2004; Erkkila & Lichtenstein, 2006). General lifestyle advice of regular physical exercise and a diverse diet with a high variety of fruits and vegetables (Ho et al., 2001; Bokura & Kobayashi, 2003; Dharmarajan et al., 2003; Ni et al., 2004; Pereira et al., 2004) supports

the intake of fibre-rich foods. These foods are also richer in vitamins, minerals and antioxidants in general, and low in cholesterol and saturated fat (Brown et al., 1999; Ahmed et al., 2014).

The amount of fibre currently needed to effectively reduce cholesterol is high compared to statins. These high levels are generally acceptable for food, but might impair actual use as treatment. Before fibres can gain a place in therapy against hypercholesterolaemia, they need to become more effective. At present, optimisation of existing fibres (changing the molecular weight, viscosity, degree of hydrophobic substitution) has not yet led to a fibre that is active enough to be comparable to statin treatment. However, this approach holds significant promise for future research. Several studies suggest that the cholesterol-modulating effect can be enhanced by combining several fibres (Gallaher et al., 2002; Parolini et al., 2013), because every fibre has a unique profile of mechanisms and effects (Carabin et al., 2009; Reimer et al., 2013).

Of the three studied fibres, the effect of HPMC is most dependent on viscosity. It has been shown that the effectiveness of pectin is also related to its molecular shape and the degree of esterification. In the case of chitosan, the molecular charge adds to the cholesterol-modulating effect. Using these facts, it should be possible to design a dietary fibre that yields an optimal cholesterol-lowering effect, using experiences in tailoring physicochemical properties and exploiting the primarily biophysical mechanisms of action. The optimal fibre should be highly viscous and have side chains with a positive charge, so they allow for adsorption of cholesterol and bile acids. It should also be stable in a gastro-intestinal environment and not be too easily digested by bacteria, since digestion would decrease adsorption and faecal excretion of bile acids and cholesterol, and lead to side effects. Fermentation of pectin is thought to be responsible for side effects like abdominal cramping and flatulence (Maki et al., 2000), and the explanation for the better tolerability of HPMC, which is not fermentable (Reppas et al., 2009; Yokoyama et al., 2011; Maki et al., 2000; Carr et al., 2003).

For now the challenge is how to package a sufficient fibre load in an acceptable and palatable way in designer foods. For the general population, adding a mix of selected fibres to products used daily like noodles (Jitpukdeebodintra & Jangwang, 2009), bread (Ausar et al., 2003) or orange juice (Maki et al., 2000) could be an option.

In conclusion, soluble dietary fibres are a promising tool for cholesterol reduction. However, fibres that are currently being used are not adequately potent in typical pharmacological dosages and need optimisation and/or to be combined with other substances.

6.2.6.1 Acknowledgements

The authors would like to thank Julia Challinor and Cassandra Nemzoff for their English manuscript correction services.

6.2.6.2 Author contribution

TvdG, AH, CvH, HP and TP designed the research, TvdG, AH and CvH provided essential materials, conducted the research, analysed data and performed the statistical analysis. TvdG, AH, and TP wrote the paper. TvdG, AH, HP and TP had primary responsibility for final content. All authors have read and approved the final manuscript.

6.2.6.3 Funding

Regular institutional funding.

6.2.7 Reference List

Ahmed, F. A., Abdel-Lattife, M. S., Abd-El Azeem, A. S., Hegazy, A. M., Hassouna, H. Z., & Algalaly, M. A. (2014). The role of chitosan and wheat germ as antidiabetic substances in diabetic rats. Research Journal of Pharmaceutical, Biological and Chemical Sciences, 5, 457-469.

Aller, R., De Luis, D. A., Izaola, O., La, C. F., Del, O. L., Fernandez, L. et al. (2004). Effect of soluble fiber intake in lipid and glucose leves in healthy subjects: A randomized clinical trial. Diabetes Research and Clinical Practice, 65, 7-11.

Anandan, R., Ganesan, B., Obulesu, T., Mathew, S., Asha, K. K., Lakshmanan, P. T. et al. (2013). Antiaging effect of dietary chitosan supplementation on glutathione-dependent antioxidant system in young and aged rats. Cell Stress Chaperones., 18, 121-125.

Anderson, J. W., Baird, P., Davis, J., Ferreri, S., Knudtson, M., Koraym, A. et al. (2009). Health benefits of dietary fiber. Nutrition Reviews, 67, 188-205.

Anraku, M., Fujii, T., Furutani, N., Kadowaki, D., Maruyama, T., Otagiri, M. et al. (2009). Antioxidant effects of a dietary supplement: reduction of indices of oxidative stress in normal subjects by water-soluble chitosan. Food Chem.Toxicol., 47, 104-109.

Anraku, M., Michihara, A., Yasufuku, T., Akasaki, K., Tsuchiya, D., Nishio, H. et al. (2010). The antioxidative and antilipidemic effects of different molecular weight chitosans in metabolic syndrome model rats. Biol.Pharm.Bull., 33, 1994-1998.

Aprikian, O., Duclos, V., Guyot, S., Besson, C., Manach, C., Bernalier, A. et al. (2003). Apple pectin and a polyphenol-rich apple concentrate are more

effective together than separately on cecal fermentations and plasma lipids in rats. Journal of Nutrition, 133, 1860-1865.

Asha, K. K. & Nair, P. G. V. (2005). Effect of chitin-chitosan treatment on fatty liver in rats with a high fat diet. Journal of Clinical Biochemistry and Nutrition, 37, 87-94.

Ausar, S. F., Morcillo, M., Leon, A. E., Ribotta, P. D., Masih, R., Vilaro, M. M. et al. (2003). Improvement of HDL- and LDL-cholesterol levels in diabetic subjects by feeding bread containing chitosan. J.Med.Food, 6, 397-399.

Baker, W. L., Tercius, A., Anglade, M., White, C. M., & Coleman, C. I. (2009). A meta-analysis evaluating the impact of chitosan on serum lipids in hypercholesterolemic patients. Ann.Nutr.Metab, 55, 368-374.

Ban, S. J., Rico, C. W., Um, I. C., & Kang, M. Y. (2012). Comparative evaluation of the hypolipidemic effects of hydroxyethyl methylcellulose (HEMC) and hydroxypropyl methylcellulose (HPMC) in high fat-fed mice. Food and Chemical Toxicology, 50, 130-134.

Bartley, G. E., Yokoyama, W., Young, S. A., Anderson, W. H., Hung, S. C., Albers, D. R. et al. (2010). Hypocholesterolemic effects of hydroxypropyl methylcellulose are mediated by altered gene expression in hepatic bile and cholesterol pathways of male hamsters. Journal of Nutrition, 140, 1255-1260.

Bazzano, L. A. (2008). Effects of soluble dietary fiber on low-density lipoprotein cholesterol and coronary heart disease risk. Current Atherosclerosis Reports, 10, 473-477.

Bergen, W. G. & Mersmann, H. J. (2005). Comparative aspects of lipid metabolism: impact on contemporary research and use of animal models. J.Nutr., 135, 2499-2502.

Bokura, H. & Kobayashi, S. (2003). Chitosan decreases total cholesterol in women: a randomized, double-blind, placebo-controlled trial. Eur.J.Clin.Nutr., 57, 721-725.

Brockman, D. A., Chen, X., & Gallaher, D. D. (2012). Hydroxypropyl methylcellulose, a viscous solublefiber, reduces insulin resistance and decreases fattyliver in Zucker Diabetic Fatty rats. Nutrition & Metabolism, 100.

Brouns, F., Theuwissen, E., Adam, A., Bell, M., Berger, A., & Mensink, R. P. (2012). Cholesterol-lowering properties of different pectin types in mildly hyper-cholesterolemic men and women. European Journal of Clinical Nutrition, 66, 591-599.

Brown, L., Rosner, B., Willett, W. W., & Sacks, F. M. (1999). Cholesterol-lowering effects of dietary fiber: a meta-analysis. Am.J Clin.Nutr, 69, 30-42.

Brufau, G., Canela, M. A., & Rafecas, M. (2007). Phytosterols, but not pectin, added to a high-saturated-fat diet modify saturated fatty acid excretion in relation to chain length. Journal of Nutritional Biochemistry, 18, 580-586.

Burdock, G. A. (2007). Safety assessment of hydroxypropyl methylcellulose as a food ingredient. Food Chem.Toxicol., 45, 2341-2351.

Carabin, I. G., Lyon, M. R., Wood, S., Pelletier, X., Donazzolo, Y., & Burdock, G. A. (2009). Supplementation of the diet with the functional fiber PolyGlycoplex is well tolerated by healthy subjects in a clinical trial. Nutr J, 8, 9.

Carr, T. P., Wood, K. J., Hassel, C. A., Bahl, R., & Gallaher, D. D. (2003). Raising intestinal contents viscosity leads to greater excretion of neutral steroids but not bile acids in hamsters and rats. Nutrition Research, 23, 91-102.

Chang, H.-P., Yao, H.-T., & Chiang, M.-T. (2012). Effects of high and low molecular weight chitosan on plasma cholesterol, glucose and adipocytokines in diabetic rats induced by streptozotocin and nicotinamide. Journal of Food and Drug Analysis, 20, 661-667+716.

Chiang, M. T., Yao, H. T., & Chen, H. C. (2000). Effect of dietary chitosans with different viscosity on plasma lipids and lipid peroxidation in rats fed on a diet enriched with cholesterol. Biosci.Biotechnol.Biochem., 64, 965-971.

Choi, C. R., Kim, E. K., Kim, Y. S., Je, J. Y., An, S. H., Lee, J. D. et al. (2012). Chitooligosaccharides decreases plasma lipid levels in healthy men. Int.J.Food Sci Nutr., 63, 103-106.

Cos, E., Ramjiganesh, T., Roy, S., Yoganathan, S., Nicolosi, R. J., & Fernandez, M. L. (2001). Soluble fiber and soybean protein reduce atherosclerotic lesions in guinea pigs. Sex and hormonal status determine lesion extension. Lipids, 36, 1209-1216.

Cox, L. M., Cho, I., Young, S. A., Anderson, W. H., Waters, B. J., Hung, S. C. et al. (2013). The nonfermentable dietary fiber hydroxypropyl methylcellulose modulates intestinal microbiota. The FASEB Journal, 27, 692-702.

Dakhara, S. L., Anajwala, C. C., & Selote, V. S. (2012). Fibrous drugs for curing various common health problems. Pharmacognosy Reviews, 6, 16-21.

Department of Agriculture., National Agricultural Library., National Academy of Sciences., Institute of Medicine., & Food and Nutrition Board. (2005). Dietary Reference Intakes for Energy, Carbohydrate, fibre, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids (Macronutrients) (2005) Chapter 7: Dietary, Functional and Total fibre.

Dharmarajan, T. S., Ravunniarath, J., & Pitchumoni, C. S. (2003). Dietary fiber: Its role in older adults. Practical Gastroenterology, 27, 43-44+46.

Dongowski, G. & Lorenz, A. (2004). Intestinal steroids in rats are influenced by the structural parameters of pectin. Journal of Nutritional Biochemistry, 15, 196-205.

Dvir, I., Chayoth, R., Sod-Moriah, U., Shany, S., Nyska, A., Stark, A. H. et al. (2000). Soluble polysaccharide and biomass of red microalga Porphyridium sp.

alter intestinal morphology and reduce serum cholesterol in rats. British Journal of Nutrition, 84, 469-476.

EFSA Panel on Dietetic Products, N. a. A. N. (2010a). Scientific Opinion on Dietary Reference Values for carbohydrates and dietary fibre. EFSA Journal, 8(3):1462.

EFSA Panel on Dietetic Products, N. a. A. N. (2010b). Scientific Opinion on the substantiation of health claims related to pectins and reduction of post-prandial glycaemic responses (ID 786), maintenance of normal blood cholesterol concentrations (ID 818) and increase in satiety leading to a reduction in energy intake (ID 4692) pursuant to Article 13(1) of Regulation (EC) No 1924/2006. EFSA Journal, 2010;8, 1747.

Erkkila, A. T. & Lichtenstein, A. H. (2006). Fiber and cardiovascular disease risk: how strong is the evidence? The Journal of cardiovascular nursing, 21, 3-8.

Eussen, S., Klungel, O., Garssen, J., Verhagen, H., Van, K. H., Van, L. H. et al. (2010). Support of drug therapy using functional foods and dietary supplements: Focus on statin therapy. British Journal of Nutrition, 103, 1260-1277.

Fernandez, M. L., Abdel-Fattah, G., & McNamara, D. J. (1993). Dietary fat saturation modifies the metabolism of LDL subfractions in guinea pigs. Arterioscler.Thromb., 13, 1418-1428.

Fernandez, M. L., Wilson, T. A., Conde, K., Vergara-Jimenez, M., & Nicolosi, R. J. (1999). Hamsters and guinea pigs differ in their plasma lipoprotein cholesterol distribution when fed diets varying in animal protein, soluble fiber, or cholesterol content. J.Nutr., 129, 1323-1332.

Fernandez, M. L. & West, K. L. (2005). Mechanisms by which Dietary Fatty Acids Modulate Plasma Lipids1. The Journal of Nutrition, 135, 2075-2078.

Gallaher, C. M. & Gallaher, D. D. (2009). Dried plums (prunes) reduce atherosclerosis lesion area in apolipoprotein E-deficient mice. British Journal of Nutrition, 101, 233-239.

Gallaher, C. M., Munion, J., Hesslink, R., Jr., Wise, J., & Gallaher, D. D. (2000). Cholesterol reduction by glucomannan and chitosan is mediated by changes in cholesterol absorption and bile acid and fat excretion in rats. J.Nutr., 130, 2753-2759.

Gallaher, D. D., Gallaher, C. M., Mahrt, G. J., Carr, T. P., Hollingshead, C. H., Hesslink, R., Jr. et al. (2002). A glucomannan and chitosan fiber supplement decreases plasma cholesterol and increases cholesterol excretion in overweight normocholesterolemic humans. J.Am.Coll.Nutr., 21, 428-433.

Gazi, I. F., Tsimihodimos, V., Tselepis, A. D., Elisaf, M., & Mikhailidis, D. P. (2007). Clinical importance and therapeutic modulation of small dense low-density lipoprotein particles. Expert.Opin.Biol.Ther., 7, 53-72.

Grizard, D., Dalle, M., & Barthomeuf, C. (2001). Changes in insulin and corticosterone levels may partly mediate the hypolipidemic effect of

guar gum and low-molecular weight pectin in rats. Nutrition Research, 21, 1185-1190.

Guha, S., Pal, S. K., Chatterjee, N., Sarkar, G., Pal, S., Guha, S. et al. (2005). Effect of chitosan on lipid levels when administered concurrently with atorvastatin--a placebo controlled study. J.Indian Med.Assoc., 103, 418, 420.

Gunness, P. & Gidley, M. J. (2010). Mechanisms underlying the cholesterol-lowering properties of soluble dietary fibre polysaccharides. Food & function, 1, 149-155.

Hernandez-Gonzalez, S. O., Gonzalez-Ortiz, M., Martinez-Abundis, E., & Robles-Cervantes, J. A. (2010). Chitosan improves insulin sensitivity as determined by the euglycemic-hyperinsulinemic clamp technique in obese subjects. Nutr.Res., 30, 392-395.

Ho, S. C., Tai, E. S., Eng, P. H., Tan, C. E., & Fok, A. C. (2001). In the absence of dietary surveillance, chitosan does not reduce plasma lipids or obesity in hypercholesterolaemic obese Asian subjects. Singapore Med.J., 42, 6-10.

Hong, Y. J., Turowski, M., Lin, J. T., & Yokoyama, W. H. (2007). Simultaneous characterization of bile acid, sterols, and determination of acylglycerides in feces from soluble cellulose-fed hamsters using HPLC with evaporative light-scattering detection and APCI-MS. Journal of Agricultural and Food Chemistry, 55, 9750-9757.

Hossain, S., Rahman, A., Kabir, Y., Shams, A. A., Afros, F., & Hashimoto, M. (2007). Effects of shrimp (Macrobracium rosenbergii)-derived chitosan on plasma lipid profile and liver lipid peroxide levels in normo- and hypercholesterolaemic rats. Clin.Exp.Pharmacol.Physiol, 34, 170-176.

Hsieh, Y. L., Yao, H. T., Cheng, R. S., & Chiang, M. T. (2012). Chitosan reduces plasma adipocytokines and lipid accumulation in liver and adipose tissues and ameliorates insulin resistance in diabetic rats. J.Med.Food, 15, 453-460.

Hung, S. C., Anderson, W. H., Albers, D. R., Langhorst, M. L., & Young, S. A. (2011). Effect of hydroxypropyl methylcellulose on obesity and glucose metabolism in a diet-induced obesity mouse model. J.Diabetes, 3, 158-167.

Hung, S. C., Bartley, G., Young, S. A., Albers, D. R., Dielman, D. R., Anderson, W. H. et al. (2009). Dietary fiber improves lipid homeostasis and modulates adipocytokines in hamsters. J.Diabetes, 1, 194-206.

Hur, S. J., Lee, S. Y., & Lee, S. J. (2015). Effect of biopolymer encapsulation on the digestibility of lipid and cholesterol oxidation products in beef during in vitro human digestion. Food Chem., 166, 254-260.

Jaffer, S. & Sampalis, J. S. (2007). Efficacy and safety of chitosan HEP-40 in the management of hypercholesterolemia: a randomized, multicenter, placebo-controlled trial. Altern.Med.Rev., 12, 265-273.

Jitpukdeebodintra, S. & Jangwang, A. (2009). Instant noodles with pectin for weight reduction. Journal of Food, Agriculture and Environment, 7, 126-129.

Jokinen, M. P., Clarkson, T. B., & Prichard, R. W. (1985). Animal models in atherosclerosis research. Exp.Mol.Pathol., 42, 1-28.

Jull, A. B., Ni, M. C., Bennett, D. A., Dunshea-Mooij, C. A., & Rodgers, A. (2008). Chitosan for overweight or obesity. Cochrane.Database.Syst.Rev., CD003892.

Jun, S. C., Jung, E. Y., Hong, Y. H., Park, Y., Kang, D. H., Chang, U. J. et al. (2012). Anti-obesity effects of chitosan and psyllium husk with L-ascorbic acid in Guinea pigs. International Journal for Vitamin and Nutrition Research, 82, 113-120.

Kim, H., Bartley, G. E., Young, S. A., Seo, K. H., & Yokoyama, W. (2013). Altered hepatic gene expression profiles associated with improved fatty liver, insulin resistance, and intestinal permeability after hydroxypropyl methylcellulose (HPMC) supplementation in diet-induced obese mice. Journal of Agricultural and Food Chemistry, 61, 6404-6411.

Kim, H., Turowski, M., Anderson, W. H., Young, S. A., Kim, Y., & Yokoyama, W. (2011). Supplementation of hydroxypropyl methylcellulose into yeast leavened all-whole grain barley bread potentiates cholesterol-lowering effect. Journal of Agricultural and Food Chemistry, 59, 7672-7678.

Kim, J.-G., Jo, S.-H., Ha, K.-S., Kim, S.-C., Kim, Y.-C., Apostolidis, E. et al. (2014). Effect of long-term supplementation of low molecular weight chitosan oligosaccharide (GO2KA1) on fasting blood glucose and HbA1c in db/db mice model and elucidation of mechanism of action. BMC Complementary and Alternative Medicine, 14.

Kim, S.-K., Park, P.-J., Yang, H.-P., & Han, S.-S. (2001). Subacute toxicity of chitosan oligosaccharide in Sprague-Dawley rats. Arzneimittel-Forschung/Drug Research, 51, 769-774.

Kirat, D. (2010). Effect of pectin feeding on monocarboxylate transporters in rat adrenal gland. Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology, 180, 57-65.

Krzysik, M., Grajeta, H., Prescha, A., & Weber, R. (2011). Effect of cellulose, pectin and chromium(III) on lipid and carbohydrate metabolism in rats. Journal of Trace Elements in Medicine and Biology, 25, 97-102.

Lau, J., Ioannidis, J. P., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. BMJ, 333, 597-600.

Lee, H. W., Park, Y. S., Choi, J. W., Yi, S. Y., & Shin, W. S. (2003). Antidiabetic effects of chitosan oligosaccharides in neonatal streptozotocin-induced noninsulin-dependent diabetes mellitus in rats. Biol.Pharm.Bull., 26, 1100-1103.

Liu, X., Yang, F., Song, T., Zeng, A., Wang, Q., Sun, Z. et al. (2012). Therapeutic effect of carboxymethylated and quanternized chitosan on insulin

resistance in high-fat-diet-induced rats and 3T3-L1 adipocytes. Journal of Biomaterials Science, Polymer Edition, 23, 1271-1284.

Luo, P., Zhang, C.-J., Feng, C., & Chen, D.-Y. (2004). Effect of compound chitosan on body mass and blood lipid level of rats in growing period. Chinese Journal of Clinical Rehabilitation, 8, 5977-5979.

Maki, K. C., Carson, M. L., Kerr Anderson, W. H., Geohas, J., Reeves, M. S., Farmer, M. V. et al. (2009a). Lipid-altering effects of different formulations of hydroxypropylmethylcellulose. Journal of Clinical Lipidology, 3, 159-166.

Maki, K. C., Carson, M. L., Miller, M. P., Anderson, W. H., Turowski, M., Reeves, M. S. et al. (2009b). Hydroxypropylmethylcellulose lowers cholesterol in statin-treated men and women with primary hypercholesterolemia. European Journal of Clinical Nutrition, 63, 1001-1007.

Maki, K. C., Carson, M. L., Miller, M. P., Turowski, M., Bell, M., Wilder, D. M. et al. (2007). High-viscosity hydroxypropylmethylcellulose blunts postprandial glucose and insulin responses. Diabetes Care, 30, 1039-1043.

Maki, K. C., Davidson, M. H., Torri, S., Ingram, K. A., O'Mullane, J., Daggy, B. P. et al. (2000). High-molecular-weight hydroxypropylmethylcellulose taken with or between meals is hypocholesterolemic in adult men. Journal of Nutrition, 130, 1705-1710.

Maki, K. C., Reeves, M. S., Carson, M. L., Miller, M. P., Turowski, M., Rains, T. M. et al. (2009c). Dose-response characteristics of high-viscosity hydroxypropylmethylcellulose in subjects at risk for the development of type 2 diabetes mellitus. Diabetes Technol.Ther., 11, 119-125.

Marounek, M., Volek, Z., Duskova, D., Tuma, J., & Taubner, T. (2013). Dose-response efficacy and long-term effect of the hypocholesterolemic effect of octadecylpectinamide in rats. Carbohydrate Polymers, 97, 772-775.

Marounek, M., Volek, Z., Synytsya, A., & Copikova, J. (2007). Effect of pectin and amidated pectin on cholesterol homeostasis and cecal metabolism in rats fed a high-cholesterol diet. Physiological Research, 56, 433-442.

Metzger, B. T., Barnes, D. M., & Reed, J. D. (2009). A comparison of pectin, polyphenols, and phytosterols, alone or in combination, to lovastatin for reduction of serum lipids in familial hypercholesterolemic swine. Journal of Medicinal Food, 12, 854-860.

Mikhailidis, D. P., Elisaf, M., Rizzo, M., Berneis, K., Griffin, B., Zambon, A. et al. (2011). European panel on low density lipoprotein (LDL) subclasses: a statement on the pathophysiology, atherogenicity and clinical significance of LDL subclasses: executive summary. Curr.Vasc.Pharmacol., 9, 531-532.

Ni, M. C., Poppitt, S. D., McGill, A.-T., Leahy, F. E., Bennett, D. A., Lin, R. B. et al. (2004). The effect of the dietary supplement, Chitosan, on body weight: A randomised controlled trial in 250 overweight and obese adults. International Journal of Obesity, 28, 1149-1156.

Nordestgaard, B. G. & Varbo, A. (2014). Triglycerides and cardiovascular disease. Lancet, 384, 626-635.

Park, H. S., Choi, J. S., & Kim, K. H. (2000). Docosahexaenoic acid-rich fish oil and pectin have a hypolipidemic effect, but pectin increases risk factor for colon cancer in rats. Nutrition Research, 20, 1783-1794.

Parolini, C., Manzini, S., Busnelli, M., Rigamonti, E., Marchesi, M., Diani, E. et al. (2013). Effect of the combinations between pea proteins and soluble fibres on cholesterolaemia and cholesterol metabolism in rats. Br.J.Nutr., 110, 1394-1401.

Pereira, M. A., O'Reilly, E., Augustsson, K., Fraser, G. E., Goldbourt, U., Heitmann, B. L. et al. (2004). Dietary fiber and risk of coronary heart disease: a pooled analysis of cohort studies. Arch.Intern.Med, 164, 370-376.

Rader, D. J. & Hovingh, G. K. (2014). HDL and cardiovascular disease. Lancet, 384, 618-625.

Reimer, R. A., Yamaguchi, H., Eller, L. K., Lyon, M. R., Gahler, R. J., Kacinik, V. et al. (2013). Changes in visceral adiposity and serum cholesterol with a novel viscous polysaccharide in Japanese adults with abdominal obesity. Obesity (Silver.Spring), 21, E379-E387.

Reppas, C., Swidan, S. Z., Tobey, S. W., Turowski, M., & Dressman, J. B. (2009). Hydroxypropylmethylcellulose significantly lowers blood cholesterol in mildly hypercholesterolemic human subjects. European Journal of Clinical Nutrition, 63, 71-77.

Rizzo, M., Giglio, R. V., Nikolic, D., Patti, A. M., Campanella, C., Cocchi, M. et al. (2013). Effects of Chitosan on Plasma Lipids and Lipoproteins: A 4-Month Prospective Pilot Study. Angiology, 65, 538-542.

Robins, S. J., Fasulo, J. M., Patton, G. M., Schaefer, E. J., Smith, D. E., & Ordovas, J. M. (1995). Gender differences in the development of hyperlipemia and atherosclerosis in hybrid hamsters. Metabolism, 44, 1326-1331.

Sanchez, D., Muguerza, B., Moulay, L., Hernandez, R., Miguel, M., & Aleixandre, A. (2008). Highly methoxylated pectin improves insulin resistance and other cardiometabolic risk factors in Zucker fatty rats. Journal of Agricultural and Food Chemistry, 56, 3574-3581.

Santas, J., Espadaler, J., Mancebo, R., & Rafecas, M. (2012). Selective in vivo effect of chitosan on fatty acid, neutral sterol and bile acid excretion: a longitudinal study. Food Chem., 134, 940-947.

Schwab, U., Louheranta, A., Torronen, A., & Uusitupa, M. (2006). Impact of sugar beet pectin and polydextrose on fasting and postprandial glycemia and fasting concentrations of serum total and lipoprotein lipids in middle-aged subjects with abnormal glucose metabolism. European Journal of Clinical Nutrition, 60, 1073-1080.

Spady, D. K. & Dietschy, J. M. (1983). Sterol synthesis in vivo in 18 tissues of the squirrel monkey, guinea pig, rabbit, hamster, and rat. J.Lipid Res., 24, 303-315.

Sumiyoshi, M. & Kimura, Y. (2006). Low molecular weight chitosan inhibits obesity induced by feeding a high-fat diet long-term in mice. J.Pharm.Pharmacol., 58, 201-207.

Tai, T.-S., Sheu, W. H. H., Lee, W.-J., Yao, H.-T., & Chiang, M.-T. (2000). Effect of chitosan on plasma lipoprotein concentrations in type 2 diabetic subjects with hypercholesterolemia [2]. Diabetes Care, 23, 1703-1704.

Tang, Z. R., Yin, Y. L., Nyachoti, C. M., Huang, R. L., Li, T. J., Yang, C. et al. (2005). Effect of dietary supplementation of chitosan and galacto-mannan-oligosaccharide on serum parameters and the insulin-like growth factor-I mRNA expression in early-weaned piglets. Domest.Anim Endocrinol., 28, 430-441.

Terpstra, A. H., Lapre, J. A., De Vries, H. T., & Beynen, A. C. (2002a). Intact pectin and its polygalacturonic acid regions have similar hypocholesterolemic properties in hybrid F1B hamsters. Die Nahrung, 46, 83-86.

Terpstra, A. H. M., Lapre, J. A., De Vries, H. T., & Beynen, A. C. (2002b). The hypocholesterolemic effect of lemon peels, lemon pectin, and the waste stream material of lemon peels in hybrid F1B hamsters. European Journal of Nutrition, 41, 19-26.

Thakur, B. R., Singh, R. K., & Handa, A. K. (1997). Chemistry and uses of pectin--a review. Crit Rev.Food Sci.Nutr., 37, 47-73.

Tharanathan, R. N. & Kittur, F. S. (2003). Chitin--the undisputed biomolecule of great potential. Crit Rev.Food Sci.Nutr., 43, 61-87.

Theuwissen, E. & Mensink, R. P. (2008). Water-soluble dietary fibers and cardiovascular disease. Physiol Behav., 94, 285-292.

van Bennekum, A. M., Nguyen, D. V., Schulthess, G., Hauser, H., & Phillips, M. C. (2005). Mechanisms of cholesterol-lowering effects of dietary insoluble fibres: relationships with intestinal and hepatic cholesterol parameters. Br.J.Nutr., 94, 331-337.

Vitic, J. & Stevanovic, J. (1993). Comparative studies of the serum lipoproteins and lipids in some domestic, laboratory and wild animals. Comp Biochem.Physiol B, 106, 223-229.

Wang, D., Han, J., Yu, Y., Li, X., Wang, Y., Tian, H. et al. (2011). Chitosan oligosaccharide decreases very-low-density lipoprotein triglyceride and increases high-density lipoprotein cholesterol in high-fat-diet-fed rats. Exp.Biol.Med. (Maywood.), 236, 1064-1069.

Xiangdong, L., Yuanwu, L., Hua, Z., Liming, R., Qiuyan, L., & Ning, L. (2011). Animal models for the atherosclerosis research: a review. Protein Cell, 2, 189-201.

Yao, H. T. & Chiang, M. T. (2002). Plasma lipoprotein cholesterol in rats fed a diet enriched in chitosan and cholesterol. J.Nutr.Sci Vitaminol. (Tokyo), 48, 379-383.

Yao, H. T. & Chiang, M. T. (2006a). Chitosan shifts the fermentation site toward the distal colon and increases the fecal short-chain fatty acids

concentrations in rats. International Journal for Vitamin and Nutrition Research, 76, 57-64.

Yao, H. T., Huang, S. Y., & Chiang, M. T. (2008). A comparative study on hypoglycemic and hypocholesterolemic effects of high and low molecular weight chitosan in streptozotocin-induced diabetic rats. Food Chem.Toxicol., 46, 1525-1534.

Yao, H.-T. & Chiang, M.-T. (2006b). Effect of chitosan on plasma lipids, hepatic lipids, and fecal bile acid in hamsters. Journal of Food and Drug Analysis, 14, 183-189.

Ylitalo, R., Lehtinen, S., Wuolijoki, E., Ylitalo, P., & Lehtimaki, T. (2002). Cholesterol-lowering properties and safety of chitosan. Arzneimittelforschung., 52, 1-7.

Yokoyama, W., Anderson, W. H. K., Albers, D. R., Hong, Y.-J., Langhorst, M. L., Hung, S.-C. et al. (2011). Dietary hydroxypropyl methylcellulose increases excretion of saturated and trans fats by hamsters fed fast food diets. Journal of Agricultural and Food Chemistry, 59, 11249-11254.

Zhang, H. L., Zhong, X. B., Tao, Y., Wu, S. H., & Su, Z. Q. (2012). Effects of chitosan and water-soluble chitosan micro- and nanoparticles in obese rats fed a high-fat diet. Int.J.Nanomedicine., 7, 4069-4076.

Zhang, J., Liu, J., Li, L., & Xia, W. (2008). Dietary chitosan improves hypercholesterolemia in rats fed high-fat diets. Nutr.Res., 28, 383-390.

Zhang, J., Zhang, W., Mamadouba, B., & Xia, W. (2012). A comparative study on hypolipidemic activities of high and low molecular weight chitosan in rats. Int.J.Biol.Macromol., 51, 504-508.

Zong, C., Yu, Y., Song, G., Luo, T., Li, L., Wang, X. et al. (2012). Chitosan oligosaccharides promote reverse cholesterol transport and expression of scavenger receptor BI and CYP7A1 in mice. Exp.Biol.Med. (Maywood.), 237, 194-200.

6.3 Evaluation

6.3.1 Strengths

The strength of this paper is the integration of findings in many different trials, so that an overall effect can be found with much more confidence than any of the individual trials could generate. It also explored the mechanisms behind the measured effects, so that the beneficial properties of specific fibres can be combined to create one designer fibre.

Similar to the gene doping SR, the selection of these three fibres was based on a consensus and informal literature search, so a fibre that is more appropriate to research might exist. It is not likely that we missed such fibres that have been researched as thoroughly as these, but there may have been more fibres that would qualify by our criteria and would have added another mechanism.

6.3.2 Limitations

The main weakness of this study is that it is based on animal research. In rodents, most cholesterol is carried in HDL, whereas humans have more LDL. Also, rats, mice, hamsters, guinea pigs and swine are used for this study, making it a heterogeneous selection. The average results collate results from different species, so is meaningless in practice.

Furthermore, the graphs showed adjusted data, adjusted for the fraction of the diet that was fibre. This allowed us to compare cholesterol changes across studies with different levels of fibres in the diets. This does assume a linear relationship between fibre content and cholesterol changes, which is not likely to be true.

Due to the lack of a reliable registry of trials in animals, it was not possible to estimate non-reporting bias or publication bias that way. The Cochrane collaboration recommends using funnel plots instead in this situation, so we compiled one.

The funnel plot was based on studies in different species, which might confound the effect. It might be that large studies use different animals (e.g., smaller animals that are easier to handle, such as mice) than small studies, and in those animals the fibres have a different effect size, so the gap at the negative outcomes cannot reliably be attributed to publication bias alone.

Displaying the funnel plot with all three fibres allows for a more reliable estimation of publication bias, as there are enough studies to create a meaningful plot and interpretation. This does assume a class effect, i.e., all studies with these three fibres in animals are subject to a publication bias of similar magnitude, which is a reasonable assumption. If a larger dataset were available, it would have been more reliable to produce three individual plots and assess the bias per fibre or even per species.

For a similar reason, we did not compile a forest plot, such as many SRs with MA would have done. The selected studies were not directly comparable, due to the use of different species, so displaying them would make it seem like they were. Also, for many studies the confidence intervals were not reported. Even our primary outcome, changes in each cholesterol level, was not consistently reported. Separate forest plots for LDL, HDL, VLDL and total cholesterol would have led the reader to believe in a larger and more reliable effect than is present.

6.3.3 Appropriateness of the methodology

The search methodology, of three search strings in three databases, worked well for this research goal: "to gain insight in the mechanisms that cause the cholesterol-lowering effect of selected fibres and present an overview of the effects that are seen in different animal models."

Compiling the results in graphs showed how consistent this effect was across species of animals and research method, but may have given the false impression that they are comparable. There were significant differences in methodology of the animal trials, which was not reflected in the graphs. Formal tests for heterogeneity were not possible due to the unavailability of basic data, such as standard deviations. Given the limitation in the availability of data, the methodology of searching, compiling and estimating bias was the most appropriate to find reliable results.

6.3.4 Lessons learned

If we were to repeat this study, we might try to contact authors to find standard deviations, so that we could draw a forest plot with a confidence interval for each species, if the dataset allows for that. This could then also have allowed us to perform a meta-regression analysis, to find out which factor the effects mostly depend on. We would also look

into other fibres that may have been missed now and see if sub-analyses based on species would be feasible.

This research area would lend itself well for a meta-regression, to find out if the species that are used impact the findings. Also, the time on the diet might impact the findings, so that would be a variable that could be explored with a meta-regression analysis. There are not enough studies that compare fibres head to head, so an indirect network analysis would not be feasible with this data.

7 Toward a new model of understanding, preventing and treating adolescent depression focussing on exhaustion and stress

In some research areas, even within EBM, there is a need for a breakthrough. A new insight that can accelerate research or shed new light on the pathophysiology of a disease. In that case, a systematic review is not always the most appropriate model. A narrative review, which allows for out-of-the-box thinking, is a better format for that.

For this research question, we wrote a narrative review, so that we would have the opportunity to think about a not well-understood pathophysiology and propose a new model without being backed up by what previous researchers had done.

7.1 Context

Initially, the goal of this review was to find new insights into the field of depression research, and particularly on biomarkers that might help with the definition of subpopulations and prediction of treatment response. After systematically searching the literature and reading hundreds of papers, it became clear that new answers would not be found by the integration of existing literature. The literature painted depression on a very high-level, academic scale, with limited applicability to the daily practice of a HCP.

A new approach was needed to reach a breakthrough and new insights. To reduce the impact of existing frameworks and theories on our way of thinking, we first looked at two patient cases and analysed what had happened. These cases were discussed to see what we could learn from practice, and how we could formulate a hypothesis that would improve treatment of depressed patients using the information from these cases. This led to a new framework, focussing on exhaustion, stress and coping mechanisms.

We used selected literature references to support our hypothesis from a behavioural and evolutionary perspective. Reversing the SR methodology, by first forming an opinion and then searching for matching evidence, often leads to unreliable outcomes, but helps with out-of-the-box thinking and finding a new approach to a problem.

7.2 Full text

Submitted.

van der Gronde, T., Los, L., Herremans, A., Oosting, R., Zorzanelli, R., Pieters, T.. Toward a new model of understanding, preventing and treating depression focussing on exhaustion and stress. *TBC*

Toward a new model of understanding, preventing and treating adolescent depression focussing on exhaustion and stress

Toon van der Gronde, PharmD[1], Leontien Los, MD[2], Arnoud Herremans, PhD[1], Ronald Oosting, PhD[1], Rafaela Zorzanelli, PhD[3], Toine Pieters, PhD[1]

[1] Department of Pharmaceutical Sciences, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Utrecht, the Netherlands

[2] Department of adolescent psychiatry and addiction prevention, Brijder-Jeugd, The Hague, The Netherlands.

[3] Instituto de Medicina Social, Rio de Janeiro State University, Rio de Janeiro, Brazil

Corresponding author:

Toine Pieters

PO Box 80 082, Utrecht 3584CG, The Netherlands; t.pieters@uu.nl

7.2.1 Abstract

Objective Adolescent depression is a heterogeneous disorder, with a wide variety of symptoms and inconsistent treatment response, and is not completely understood. A dysregulated stress system is a consistent finding, however, and exhaustion is a consistent trait in adolescent patients. The aim of this paper is to critically assess current hypotheses in adolescent depression research and reframe causes and treatment approaches.

Methods A mixed-method approach involved a review based on publications from Pubmed and Embase, and two exemplary adolescent cases.

Results Both cases show a spiral of stress and exhaustion, but with a different profile of symptoms and coping mechanisms. Reframing both cases from the perspective of coping behaviour, searching for the sources of experienced stress and exhaustion, showed coping similarities. This proved essential in the successful personalized treatment and recovery process. In combination with recent evidence, both cases support the functional reframing of depression as the outcome of a stress- and exhaustion-related spiralling mechanism.

Conclusions We propose moving away from a symptom-based, mood-centred view towards a model in which adolescent depression is framed as a consecutive failure of stress coping mechanisms and chronic exhaustion. Addressing exhaustion and coping primarily as a treatment strategy in adolescents and young adults might improve outcomes.

Keywords: Depression, Adolescents, Stress, Exhaustion, Treatment

7.2.2 Introduction

Major depressive disorder (MDD) is the leading cause of disability worldwide, (World Health Organisation, 2016) with 10 to 15% of patients proceeding to suicide, (Leadholm, Rothschild, Nielsen, Bech, & Ostergaard, 2014; Sharpley & Bitsika, 2013) and a substantive disease burden for adolescents and young adults. (Liu et al., 2019; Mullen, 2018; Vibhakar, Allen, Gee, & Meiser-Stedman, 2019) Depression is a heterogeneous group of brain disorders with varied contextualized origins, complex genetics and not completely understood neurobiology. The aetiology is not fully understood, and particularly for adolescents there is an evidence gap. (DeFilippis, 2018; Pedersen et al., 2019; Vibhakar et al., 2019) The serendipitous discovery of first the tri- and tetracyclic antidepressants (TCAs) and later the serotonin reuptake inhibitors (SSRIs) led successively to the catecholamine and monoamine hypotheses of depression. (Niciu, Ionescu, Richards, & Zarate, 2014) In later years, reduced adult neurogenesis and changes in structural and functional neuronal plasticity have been linked to the onset and treatment opportunities of major depression. (Akil et al., 2018; B. R. Miller & Hen, 2015) Genetic research has shown that there is not a single genetic cause for depression, and all known genetic factors combined only explain a limited percentage of the variance in clinical outcomes. (Antypa, Drago, & Serretti, 2013; Fabbri, Di Girolamo, & Serretti, 2013) The estimated heritability of depression is 35-40%, indicating 60-65% is explained by other factors, such as adverse life experiences. (Akil et al., 2018; Flint & Kendler, 2014) Researchers have turned to epigenetics to develop new forms of genetic and pharmacological modelling, in an effort to describe the aetiology of depression better. (Pieters, 2017) Despite many years of research by numerous investigators both in academia and industry, psychoactive magic bullets with controllable and specific effects on the brain microcircuitry and chemistry did not and probably will not materialize due to the complex nature of mental disorders.

(Pieters & Snelders, 2009) In order to open up our thinking about MDD we take up the challenge to reframe depression, specifically focussing on adolescents.

We see that within the current framework depression is diagnosed based on the presence of a series of mood-related symptoms and their effect on daily functioning. The seven most commonly used interviews and self-report questionnaires together describe 52 heterogenous types of symptoms, such as either high or low appetite, more or less sleep than usual, and a feeling of sadness. (Fried, 2017) This causes differences in diagnosis based on which schale is used. (Zimmerman, Walsh, Friedman, Boerescu, & Attiullah, 2017) The widely varying patterns in which these symptoms often present themselves, (Gold, Machado-Vieira, & Pavlatou, 2015; van Loo, de Jonge, Romeijn, Kessler, & Schoevers, 2012) and the high occurrence of several comorbidities, such as anxiety, psychosis and autism spectrum disorder, indicate that depression is not a homogenous disease, but a heterogeneous group of disorders associated with a wide variety of different risk factors. (DeFilippis, 2018; Ginsburg et al., 2018; Malas, Plioplys, & Pao, 2019; Vibhakar et al., 2019) This is the main driver for us to reframe the concept of MDD. It could also explain why responses to treatment vary substantially and why older age is a consistent and important risk factor for a poorer MDD course. (Fitzgerald, 2013; Harald & Gordon, 2012; Schaakxs et al., 2018; Van Loo et al., 2014) We will take a new perspective towards MDD, by focussing on stress and the depressive mood related to development in adolescence. This yields a promise for novel therapeutic approaches and potential breakthroughs in depression research, treatment and prevention.

7.2.3 Methods

A mixed method approach was used involving clinical investigation of adolescent case reports and a narrative review. Pubmed and Embase were searched for relevant publications, with select additions of recent findings based on collective suggestions of the authors. Written informed consent was obtained from both subjects for the case reports.

7.2.4 Depression and stress

Many findings in depression research have failed the scientific test of replication. For example, the volume of the amygdala of depressed patients has been found to be increased(Vassilopoulou et al., 2013) in

some studies, and decreased in others. (Jentsch et al., 2015) Patients with melancholic depression were thought to respond better to TCAs than atypical patients (hence the name), (Baldwin & Bolognesi, 2012; Gili et al., 2012; Gold et al., 2015; Harald & Gordon, 2012) but other researchers could not replicate this finding. (Arnow et al., 2015; Baumeister & Parker, 2010; Uher et al., 2011; Yang et al., 2013) Plasma levels of leptin, which reduces appetite, has been found higher in melancholic depressed patients, (Cizza et al., 2012) or higher in atypical patients. (Baune et al., 2012) Finally, childhood trauma and/or abuse is more common in melancholic than in atypical patients, (Lamers et al., 2012) or vice versa. (Li et al., 2014)

One consistent finding, however, is a dysregulated stress system in depressed patients. (Belvederi Murri et al., 2014; Fischer, Strawbridge, Vives, & Cleare, 2017; Teismann et al., 2014; Zorn et al., 2017) In approximately 70% of depressed patients a dysfunction of the HPA-axis is detected, mainly hyperactivity. (Baune et al., 2012; Holsboer, 2000) Also a disruption of the diurnal variation of cortisol is commonly seen. (Fu, Steiner, & Costafreda, 2013; Raison & Miller, 2011) Unfortunately, after years of research efforts this finding has never resulted in a stress-targeted treatment option or a clinical test to predict treatment response, (Herbert, 2013; Horstmann & Binder, 2011) and it remains debated whether HPA-axis dysregulation is a cause or a consequence of depression.

This does provide an important insight: depression is at least partially the result of stress and a differential dysregulation in the stress system is an important trait. (Baune et al., 2012) The stressor may be in the past (e.g. childhood maltreatment or trauma), (Vibhakar et al., 2019; Willner, Muscat, & Papp, 1992) or acutely (e.g. dealing with new life events). The initial response to stress is typically a coping mechanism aimed at exerting control over the stressor either by avoiding, reducing or predicting its occurrence. Examples of such efforts are the cancelling of obligations or disengagement from social interaction. (Orzechowska, Zajączkowska, Talarowska, & Gałecki, 2013) The HPA axis exerts a fundamental role regulating both internal as external stimuli, integrating the physiopathological and behavioural dimensions of stress. Depression is the result of a failure of coping mechanisms to control the stressors and a differential dysregulation in the stress system.

The accumulation of stressful events, and the eventual failure of coping mechanisms to deal with the stress, can lead to exhaustion and depressive behaviour. Preclinical experiments already hinted at a relation between the effectiveness of coping behaviour, the effort involved and feedback on the development of gastric ulcers. Although coping efforts were effective, ulcers still developed when coping took more effort and less feedback was offered. (Weiss, n.d., 1972) Preclinical evidence indicated that chronic exposure to relatively mild stressors (e.g. tilting the cage at a slight angle, emptying a water bottle in the cage, introducing new bedding material), which rats can adapt to relatively easily, ultimately resulted in the development of anhedonia. (Willner et al., 1992) The chronic character of having to cope with mild stressful events over and over again, and the lack of control over stressors, was sufficient for depressive symptoms to develop. (Weiss, 1991) We hypothesize that depressive behaviour, and specifically anhedonia and withdrawal, and consequent loss of interest and enjoyment in usual activities, is an evolutionary mechanism to guard the organism against the exhaustion that may results from excessive or chronic coping behaviour. As such, depressive behaviour is both an expression of (less-visible) psychological pressure and a physiological precaution. This substantiates the entanglement of psychological and physiological factors in MDD.

Stress response mechanisms can change the allocation of metabolic resources in a stressful situation, where that is needed. Similarly, depression could be the expression of a forced change in allocation of attention. Depressed patients are known to ruminate, or continually analyse their problems and relive their memories. Anhedonia can be interpreted as a way to secure mental resources, by reducing the interest in distractions. (Andrews & Thomson, 2009; Gold et al., 2015) Depression can be seen as an exaggerated social navigating coping mechanism, caused by an accumulation of stress and a spiral of unsuccessful adaptive behaviours which leads to exhaustion. By entering a depressive mode, the organism aims to guard itself from exhaustion. The challenge is to interfere with this mood-affecting spiralling mechanism (see fig. 1) to prevent depression from developing. Dealing with stress and potential exhaustion, as opposed to dealing with the symptoms of depression, can prove to be an effective treatment approach.

There is currently only limited evidence-based rationale for choosing one treatment over another for an individual patient, (Cipriani et al., 2018, 2016; Gili et al., 2012; Lopresti, Maker, Hood, & Drummond, 2014) with no differentiated approach for adolescents or adults. (Bernaras, Jaureguizar, & Garaigordobil, 2019; Mullen, 2018) Even defining depression subtypes based on symptoms has not helped. (Insel & Cuthbert, 2015) Despite guidelines and evidence-based interventions, treatment is still primarily based on trial and error, (Martin, Tansey, Schalkwyk, & Powell, 2015; D. B. Miller & O'Callaghan, 2013) and primarily aimed at improving mood. Yet, between one third (Rush et al., 2006) to half (Gold et al., 2015; Lichtblau, Schmidt, Schumann, Kirkby, & Himmerich, 2013) of adult patients show no response to weeks of first line treatment with antidepressant drugs, and are advised to try a different antidepressant. Further, one third of all patients never reach a response after four lines of antidepressant treatment. (Rush et al., 2006) The current therapeutic shortcomings are the consequences of our lack of knowledge of causes, the underlying neurobiology and -chemistry, and risk factors that contribute to the onset and maintenance of depression. As a consequence, the treatment paradigms are oversimplified with little attention for preventive measures. (Bockting et al., 2018)

New insights in the complex aetiology of depression might be offered by findings with the use of psychedelics for treatment-resistant depression. (Carhart-Harris et al., 2017; Roseman, Nutt, & Carhart-Harris, 2017) Several psychedelics have shown to help depressive patients in a limited number of studies with low number of patients. Many of those compounds have diverse pharmacological profiles, including robust effects on the serotonergic system. (Carhart-Harris et al., 2016; Dos Santos et al., 2016) A psychedelic not acting on the serotonergic system is ketamine, which acts as an antagonist on the N-methyl-D-aspartate (NMDA) receptor, a type of glutamate receptor. (Haile et al., 2014; Monteggia & Zarate, 2015) This highlights that serotonergic activity, or even a mono-aminergic activity, is not required for the antidepressant effect of a psychedelic compound, further stressing the need for abandoning the old hypotheses. Psychedelics and the chemically related 3,4-methylenedioxy-methamphetamine (MDMA)-assisted psychotherapy may have a place in offering a positive experience to break the self-sustaining depressive state and allowing for introspection to process stressful life-time experiences as a form of reverse medical engineering. (Wagner et al., 2017)

From a psychological point of view, psychedelics work through a different mechanism than classic antidepressants. Instead of the elevation of mood and the reduction of anxiety, psychedelic drugs induce a profound temporary positive experience (e.g. a mystical or religious sensation). This positive experience allows for the reprocessing of past emotions and introspection. Also, the use of a psychedelic in combination with a psychotherapeutic process could have long-term effects, counteracting the effect of a negative experience and disrupting the negative and 'downward spiralling' compulsive thinking. (Dos Santos et al., 2016)

In this article we move away from mood improvement as a primary target. (Müller, Myint, & Schwarz, 2011; Valkanova, Ebmeier, & Allan, 2013) We offer an alternative integrated approach for the treatment of adolescent and young adult depression by focussing on stress factors and exhaustion reduction, seeing anhedonia and withdrawal as an evolutionary coping mechanism similar to the social navigation hypothesis of Watson and Andrews. (Watson & Andrews, 2002) With this approach we take a functional perspective, and focus on the function the depressive state provides to the adolescent patient and how it develops. This perspective is instrumental for tailor-made treatment strategies.

We will discuss these insights on the basis of two adolescent patient reports. Mood disorders have been shown to be progressive, with patients developing more complex psychopathologies over time. (Fleisher & Katz, 2001; Hillegers et al., 2005) Approximately 50% of patients retrospectively state that their first depressive episode occurred before the age of 20;(Fleisher & Katz, 2001; Hawkins, 2018) another report states 50% experience that before the age of 14. (García-Carrión, Villarejo-Carballido, & Villardón-Gallego, 2019) This further highlights the progressive nature of depression and the need for early intervention.

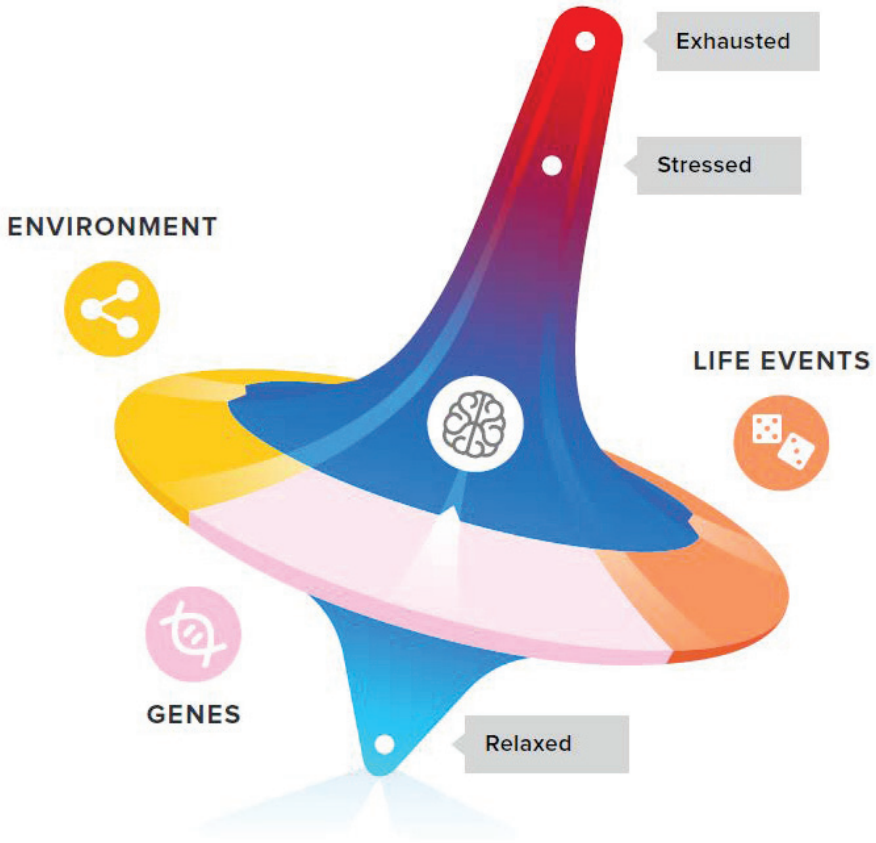

Fig. 1. Life events, genetics and environment all have an impact on the development of stress, coping with stress, and ultimately exhaustion and depression symptoms in adolescents and young adults.

7.2.5 Case reports

7.2.5.1 Case 1

A 17-year-old caucasian woman was referred by her own general practitioner to the department of adolescent psychiatry and addiction prevention for binge drinking and daily use of marijuana. The intake together with her parents showed that the patient already had a history of moderate depression and an eating disorder, anorectic of the purging type with moderate severeness. No abnormalities were reported

regarding appearance, behaviour, eye contact and rapport orientation and cognition (intelligence quotient (IQ) of 127). However, she regularly appeared to suffer from suicidal thoughts with a rather low ability to experience pleasure. She had no concrete suicide plans, in gloomy periods she showed risky behaviour, like crossing a busy road without looking. She usually performed well in school, in spite of occasional lags in attendance, which were compensated with short periods of active study. Her mother had a history of MDD.

At the department of adolescent psychiatry and addiction prevention, we classified the addiction behaviour as mild. But we also established a comorbid psychiatric and substance-use disorder profile. Thus, we chose for an integrated treatment for comorbidity that has been found to be consistently superior. (Kelly & Daley, 2013) Effective treatment for comorbid conditions combines different therapeutic modalities, i.e. psychotherapy (e.g. motivational interviewing [MI], cognitive behavioural therapy [CBT]), pharmacotherapy (e.g. antidepressants), and family therapy. Using combinations of different modalities typically increases therapeutic effect by exerting a synergistic impact on symptoms. (Mullen, 2018)

With MI, the patient was motivated to choose a first education-related treatment goal. This was to prevent school dropout at all cost. We started CBT to control her marihuana and alcohol abuse and prevent school dropout. We added medication in order to try to stabilize her mood with fluoxetine. The medication initially seemed to have some effect but after two months there was a sharp mood drop, increased suicidality and aggravation of eating disorder symptoms. Eventually she had a body mass index of 16 kg/m2. The eating problems were mapped and analysed by an eating disorder specialist. The latter used a problem-solving approach and focused on both directive counselling and emotional support. The eating disorder specialist also advised to choose a medication with low risks of weight gain. The psychiatrist changed the medication to citalopram.

Subsequently, the treatment team focused on teaching the patient how to cope with stressful situations and the associated anxiety. The stress appeared to be mainly caused by a feeling of lack of control. The patient turned out to have a high intelligence and learning ability, but also felt that she had no control over her learning process. She had not sufficiently developed social learning strategies in her early school

years. In addition, there appeared to be an issue of individuation and separation problems. These problems got worse because it was almost impossible for her parents to let her develop in her own way due to the stress they had over her suicidal thoughts, drug use and worsening physical condition due to bad eating habits. We decided on an additional family counselling approach to address these issues.

The integrated treatment modality approach proved effective. She developed a realistic idea of what caused her stress, how she reacted situationally and improved her awareness that she tends to have control over everything. Her parents were involved in helping her developing control coping skills and checking on achievements. Because of this insight, she succeeded in maintaining her diet less strictly and experimenting with behaving differently without alcohol or drugs. Her parents saw that she was doing better and were able to release her a bit more. This increased her sense of control and provided enough space to further discover what goals she wanted to achieve. In the process her mood and her ability to experience pleasure improved significantly. She successfully passed her school exams and proceeded to university.

7.2.5.2 Case 2

A 15-year old caucasian girl was referred by her own general practitioner after a suicide attempt with symptoms of sadness, anxiety and obsessive-compulsive behaviour. The intake was together with her parents. She was struggling in school, despite her very supportive family. No drug abuse or other psychiatric symptoms were found. She told the counsellor she tried hard, but felt that she could not keep up in school; it was never good enough, no matter how hard she tried. The counsellor estimated that the school level was appropriate for the level of intelligence of the patient.

She had periods when her self-esteem was very low. During these periods she spent hours on her appearance, focussing on her hair and makeup. Her hair fell out as a result of these sessions. She could not stop herself, and always ended with self-harm. This in turn lowered her self-esteem and increased the experienced stress. She was locked in a downward spiral. Gradually her mood disorder worsened and made her passive. She no longer wanted to go to school and meet friends, but passed hours in front of the mirror. She attempted to end her life.

We hypothesized on the basis of the girl's stress complaints that she felt school, parents and friends expected too much of her. After a neuropsychological assessment the testing showed that she had a disharmonic intelligence profile with an IQ of approximately 80, inconsistent at all factor levels. We classified a mild intellectual developmental disorder in the conceptual and practical domain, which explained the structural struggle with the standard school curriculum instructions. We educated parents and school on how instructions might fit in better with her learning abilities and style. Her preferred method of learning new things was being shown how to do it, as opposed to having it explained to her. This led to significant stress reduction and positive school experiences. In the process her self-esteem improved, the experienced stress decreased and her mood improved. CBT was adjusted to her learning style, and was used to reduce her obsessive-compulsive behaviour.

7.2.6 Discussion

Though both case reports show a different profile of symptoms and coping mechanisms, in both cases a downward spiral of stress and exhaustion are central. Both patients described themselves as rarely feeling relaxed and as struggling to fulfil their daily tasks. After a few years of chronic stress, a period followed in which they felt constantly exhausted.

The first patient coped with stressful situations through aberrant food intake behaviour, suicidal thoughts and mood swings and depression. She overcompensated this restrictive behaviour with recreational drug abuse. The second patient developed compulsory behaviour, stress and suicidal thoughts and overcompensated leading to self harm. In the current framework both cases would be viewed as different and based on their symptomatology ask for different treatments. Reframing both cases, from the perspective of coping behaviour, searching for the origin and sources of the experienced stress and exhaustion and coping with stressful situations, showed stress coping similarities between the two cases and proved an essential part of the personalized treatment and recovery process. Both cases support the added clinical value of the functional reframing of depression as the outcome of a mood-affecting stress and exhaustion related spiralling mechanism.

The adolescent cases presented here are good examples of how depression can be managed by relearning effective coping behaviour. This prevents patients from reverting to a depressive state in order to cope with the life stressors. In more severe and chronic cases, patients suffering from difficult to treat or treatment resistant MDD, patients are in a deep depressive state and are not capable of learning new coping behaviours. We envision in such situations that more radical medical interventions are needed to first elevate patients from the depressive state into a state where learning new and effective coping strategies can take place. In these situations, psychedelics (e.g. ketamine) have proven to be effective to temporarily draw people out of a deep depressive state. With the support of follow-up medication and adequate psychological guidance, the patients may then develop effective coping strategies.

7.2.7 Conclusion

Reframing depression and shifting clinical practice to a more comprehensive and integrated look at the individual experience of a patient, including all causes for stress, pressure and exhaustion, might be more helpful in developing promising treatment strategies. Also, shifting treatment practices towards preventive mental health interventions with a focus on stress and exhaustion, and providing coping strategies, could have a significant and lasting impact on many patients struggling with depression.

Increased focus is needed on support programs to help individuals develop functional coping mechanisms to deal with pressure, before more serious coping mechanisms develop in the form of withdrawal from stressful situations, compulsory behaviour or frequently occasional use of recreational drugs. (Hinckley & Riggs, 2019) Our intuition is that during successful treatment patients experience small successes of effective coping and re-live the rewarding properties of such experiences. Reliving experiences can repair the damaged reward mechanisms and diminishes the experienced amount of anxiety and stress which will subsequently drive and sustain further recovery. (Carhart-Harris et al., 2014) Psychedelics may offer help in facilitating introspection and re-living experiences.

Effective treatment strategies for adolescent and young adult depression should combine different therapeutic modalities and focus on exhaustion and sources of stress. Using a combination of treatment modalities could increase therapeutic effectectiveness by improving the

pace of learning new coping behaviours, exerting a synergistic impact on the developmental perspective, and breaking the downward spiral of stress and exhaustion, which eventually leads to a reduction of the depression symptoms. This might also help for other related mental disorders in adolescents and young adults where exhaustion and stress are central, such as burnout. (Humikowski, 2018) But similarly, post-traumatic stress disorder and generalised anxiety disorder are related to stress. (Okech, Hansen, Howard, Anarfi, & Burns, n.d.; Saxon et al., 2017) These disorders could also benefit from the reframing of the concept of depressive mood and stress. We would like to offer this integrated and multidisciplinary perspective as a guidance for the development of new multimodal treatment approaches for MDD and other related psychiatric disorders.

7.2.7.1 Author contributions

Conception or design of the work: TP, TvdG, LL, RO, AH, RZ

Data collection cases: LL

Data analysis and interpretation: TP, TvdG, LL, AH

Drafting the manuscript: TvdG

Critical revision of the article: TP, LL, TvdG, RO, AH, RZ

Final approval of the version to be published: TP, LL, TvdG, RO, AH, RZ

7.2.7.2 Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. All authors received regular institutional funding.

7.2.7.3 Previous presentation

This data has not been presented previously.

7.2.7.4 Patient consent

Written informed consent was obtained from the two participants for the publication of these case reports. Both participants gave their consent at an age over 16.

7.2.7.5 Acknowledgements

The authors would like to thank prof dr. Eric Vermetten and prof. dr. Claudi Bockting for their feedback on this paper, and the patients who consented to have their history described for the case reports. Furthermore, we would like to thank Cassandra Nemzoff for her English

manuscript correction services. In addition, we would like to thank Frank-Jan van Lunteren for his comprehensive graphics and the U-Talent students Emma, Frits, Imke, Hannah and Thijmen for their insightful LSD and psilocybin modelling contributions.

7.2.7.6 Conflict of interest

During the late stage development of this manuscript, TvdG accepted a position in oncology at AstraZeneca. AstraZeneca had no role in any aspect of this paper.

7.2.8 References

Akil, H., Gordon, J., Hen, R., Javitch, J., Mayberg, H., McEwen, B., … Nestler, E. J. (2018). Treatment resistant depression: A multi-scale, systems biology approach. Neuroscience and Biobehavioral Reviews, 84, 272–288. https://doi.org/10.1016/j.neubiorev.2017.08.019

Andrews, P. W., & Thomson, J. A. (2009). The bright side of being blue: depression as an adaptation for analyzing complex problems. Psychological Review, 116(3), 620–654. https://doi.org/10.1037/a0016242

Antypa, N., Drago, A., & Serretti, A. (2013). The role of COMT gene variants in depression: Bridging neuropsychological, behavioral and clinical phenotypes. Neuroscience and Biobehavioral Reviews, 37(8), 1597–1610. https://doi.org/10.1016/j.neubiorev.2013.06.006

Arnow, B. A., Blasey, C., Williams, L. M., Palmer, D. M., Rekshan, W., Schatzberg, A. F., … Rush, A. J. (2015). Depression Subtypes in Predicting Antidepressant Response: A Report From the iSPOT-D Trial. The American Journal of Psychiatry, 172(8), 743–750. https://doi.org/10.1176/appi.ajp.2015.14020181

Baldwin, D. S., & Bolognesi, F. (2012). On predicting the response to antidepressant treatment. Human Psychopharmacology, 27(4), 343–344. https://doi.org/10.1002/hup.2232

Baumeister, H., & Parker, G. (2010). A second thought on subtyping major depression. Psychotherapy and Psychosomatics, 79(6), 388–389. https://doi.org/10.1159/000320896

Baune, B., Stuart, M., Gilmour, A., Wersching, H., Heindel, W., Arolt, V., & Berger, K. (2012). The relationship between subtypes of depression and cardiovascular disease: a systematic review of biological models. Translational Psychiatry, 2(3), e92. https://doi.org/10.1038/tp.2012.18

Belvederi Murri, M., Pariante, C., Mondelli, V., Masotti, M., Atti, A. R., Mellacqua, Z., … Amore, M. (2014). HPA axis and aging in depression: systematic review and meta-analysis. Psychoneuroendocrinology, 41, 46–62. https://doi.org/10.1016/j.psyneuen.2013.12.004

Bernaras, E., Jaureguizar, J., & Garaigordobil, M. (2019). Child and Adolescent Depression: A Review of Theories, Evaluation Instruments, Prevention Programs, and Treatments. Frontiers in Psychology, 10. https://doi.org/10.3389/fpsyg.2019.00543

Bockting, C. L. H., Klein, N. S., Elgersma, H. J., van Rijsbergen, G. D., Slofstra, C., Ormel, J., … Burger, H. (2018). Effectiveness of preventive cognitive therapy while tapering antidepressants versus maintenance antidepressant treatment versus their combination in prevention of depressive relapse or recurrence (DRD study): a three-group, multicentre, randomised control. The Lancet Psychiatry, 5(5), 401–410. https://doi.org/10.1016/S2215-0366(18)30100-7

Carhart-Harris, R. L., Bolstridge, M., Rucker, J., Day, C. M. J., Erritzoe, D., Kaelen, M., … Nutt, D. J. (2016). Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. The Lancet Psychiatry, 3(7), 619–627. https://doi.org/10.1016/S2215-0366(16)30065-7

Carhart-Harris, R. L., Leech, R., Hellyer, P. J., Shanahan, M., Feilding, A., Tagliazucchi, E., … Nutt, D. (2014). The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. Frontiers in Human Neuroscience, 8, 20. https://doi.org/10.3389/fnhum.2014.00020

Carhart-Harris, R. L., Roseman, L., Bolstridge, M., Demetriou, L., Pannekoek, J. N., Wall, M. B., … Nutt, D. J. (2017). Psilocybin for treatment-resistant depression: fMRI-measured brain mechanisms. Scientific Reports, 7(1), 13187. https://doi.org/10.1038/s41598-017-13282-7

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., … Geddes, J. R. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. The Lancet. https://doi.org/10.1016/S0140-6736(17)32802-7

Cipriani, A., Zhou, X., Del Giovane, C., Hetrick, S. E., Qin, B., Whittington, C., … Xie, P. (2016). Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: a network meta-analysis. The Lancet, 388(10047), 881–890. https://doi.org/10.1016/S0140-6736(16)30385-3

Cizza, G., Ronsaville, D. S., Kleitz, H., Eskandari, F., Mistry, S., Torvik, S., … Martinez, P. E. (2012). Clinical subtypes of depression are associated with specific metabolic parameters and circadian endocrine profiles in women: The power study. PLoS ONE, 7(1). https://doi.org/10.1371/journal.pone.0028912

DeFilippis, M. (2018). Depression in Children and Adolescents with Autism Spectrum Disorder. Children, 5(9), 112. https://doi.org/10.3390/children5090112

Dos Santos, R. G., Osório, F. L., Crippa, J. A. S., Riba, J., Zuardi, A. W., & Hallak, J. E. C. (2016). Antidepressive, anxiolytic, and antiaddictive effects of ayahuasca, psilocybin and lysergic acid diethylamide (LSD): a systematic review of clinical trials published in the last 25 years. Therapeutic Advances in Psychopharmacology, 6(3), 193–213. https://doi.org/10.1177/2045125316638008

Fabbri, C., Di Girolamo, G., & Serretti, A. (2013). Pharmacogenetics of antidepressant drugs: An update after almost 20 years of research. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 162(6), 487–520. https://doi.org/10.1002/ajmg.b.32184

Fischer, S., Strawbridge, R., Vives, A. H., & Cleare, A. J. (2017). Cortisol as a predictor of psychological therapy response in depressive disorders: systematic review and meta-analysis. The British Journal of Psychiatry : The Journal of Mental Science, 210(2), 105–109. https://doi.org/10.1192/bjp.bp.115.180653

Fitzgerald, P. J. (2013). Black bile: Are elevated monoamines an etiological factor in some cases of major depression? Medical Hypotheses, 80(6), 823–826. https://doi.org/10.1016/j.mehy.2013.03.023

Fleisher, W. P., & Katz, L. Y. (2001). Early onset major depressive disorder. Paediatrics & Child Health, 6(7), 444–448. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2807758/

Flint, J., & Kendler, K. S. (2014). The Genetics of Major Depression. Neuron, 81(3), 484–503. https://doi.org/10.1016/j.neuron.2014.01.027

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. Journal of Affective Disorders, 208(July 2016), 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Fu, C. H. Y., Steiner, H., & Costafreda, S. G. (2013). Predictive neural biomarkers of clinical response in depression: A meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. Neurobiology of Disease, 52, 75–83. https://doi.org/10.1016/j.nbd.2012.05.008

García-Carrión, R., Villarejo-Carballido, B., & Villardón-Gallego, L. (2019). Children and Adolescents Mental Health: A Systematic Review of Interaction-Based Interventions in Schools and Communities. Frontiers in Psychology, 10. https://doi.org/10.3389/fpsyg.2019.00918

Gili, M., Roca, M., Armengol, S., Asensio, D., Garcia-Campayo, J., & Parker, G. (2012). Clinical Patterns and Treatment Outcome in Patients with Melancholic, Atypical and Non-Melancholic Depressions. PLoS ONE, 7(10). https://doi.org/10.1371/journal.pone.0048200

Ginsburg, A. D., Stadem, P. S., Takala, C. R., Croarkin, P. E., Mattson, A. B., Billings, M. L., … Huxsahl, J. E. (2018). An Examination of Screening Tools for Collaborative Care of Adolescent Depression. The Journal of Clinical Psychiatry, 79(4). https://doi.org/10.4088/JCP.17m11543

Gold, P. W., Machado-Vieira, R., & Pavlatou, M. G. (2015). Clinical and biochemical manifestations of depression: Relation to the neurobiology of stress. Neural Plasticity, 2015, 7–9. https://doi.org/10.1155/2015/581976

Haile, C. N., Murrough, J. W., Iosifescu, D. V, Chang, L. C., Al Jurdi, R. K., Foulkes, A., … Mathew, S. J. (2014). Plasma brain derived neurotrophic factor (BDNF) and response to ketamine in treatment-resistant depression. The International Journal of Neuropsychopharmacology, 17(02), 331–336. https://doi.org/10.1017/S1461145713001119

Harald, B., & Gordon, P. (2012). Meta-review of depressive subtyping models. Journal of Affective Disorders, 139(2), 126–140. https://doi.org/10.1016/j.jad.2011.07.015

Hawkins, E. (2018, March 3). Screen teenagers annually for depression, say US doctors. The Guardian. Retrieved from https://www.theguardian.com/society/2018/mar/03/screen-teenagers-annually-for-depression-say-us-doctors?CMP=Share_iOSApp_Other

Herbert, J. (2013). Cortisol and depression: three questions for psychiatry. Psychological Medicine, 43(3), 449–469. https://doi.org/10.1017/S0033291712000955

Hillegers, M. H., Reichart, C. G., Wals, M., Verhulst, F. C., Ormel, J., & Nolen, W. A. (2005). Five-year prospective outcome of psychopathology in the adolescent offspring of bipolar parents. Bipolar Disorders, 7(4), 344–350. https://doi.org/10.1111/j.1399-5618.2005.00215.x

Hinckley, J. D., & Riggs, P. (2019). Integrated Treatment of Adolescents with Co-occurring Depression and Substance Use Disorder. Child and Adolescent Psychiatric Clinics of North America, 28(3), 461–472. https://doi.org/10.1016/j.chc.2019.02.006

Holsboer, F. (2000). The corticosteroid receptor hypothesis of depression. Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology, 23(5), 477–501. https://doi.org/10.1016/S0893-133X(00)00159-7

Horstmann, S., & Binder, E. B. (2011). Glucocorticoids as Predictors of Treatment Response in Depression. Harvard Review of Psychiatry, 19(3), 125–143. https://doi.org/10.3109/10673229.2011.586550

Humikowski, C. A. (2018). Beyond Burnout. JAMA, 320(4), 343. https://doi.org/10.1001/jama.2018.9910

Insel, T. R., & Cuthbert, B. N. (2015). Brain disorders? Precisely. Science, 348(6234), 499–500. https://doi.org/10.1126/science.aab2358

Jentsch, M. C., Van Buel, E. M., Bosker, F. J., Gladkevich, A. V, Klein, H. C., Oude Voshaar, R. C., … Schoevers, R. A. (2015). Biomarker approaches in major depressive disorder evaluated in the context of current hypotheses. Biomarkers in Medicine, 9(3), 277–297. https://doi.org/10.2217/bmm.14.114

Kelly, T. M., & Daley, D. C. (2013). Integrated treatment of substance use and psychiatric disorders. Social Work in Public Health, 28(3–4), 388–406. https://doi.org/10.1080/19371918.2013.774673

Lamers, F., Rhebergen, D., Merikangas, K. R., de Jonge, P., Beekman, A. T. F., & Penninx, B. W. J. H. (2012). Stability and transitions of depressive subtypes over a 2-year follow-up. Psychological Medicine, 42(2012), 2083–2093. https://doi.org/10.1017/S0033291712000141

Leadholm, A. K. K., Rothschild, A. J., Nielsen, J., Bech, P., & Ostergaard, S. D. (2014). Risk factors for suicide among 34,671 patients with psychotic and non-psychotic severe depression. Journal of Affective Disorders, 156, 119–125. https://doi.org/10.1016/j.jad.2013.12.003

Li, Y., Aggen, S., Shi, S., Gao, J., Li, Y., Tao, M., … Kendler, K. S. (2014). Subtypes of major depression: latent class analysis in depressed Han Chinese women. Psychological Medicine, 44(15), 3275–3288. https://doi.org/10.1017/S0033291714000749

Lichtblau, N., Schmidt, F. M., Schumann, R., Kirkby, K. C., & Himmerich, H. (2013). Cytokines as biomarkers in depressive disorder: Current standing and prospects. International Review of Psychiatry, 25(5), 592–603. https://doi.org/10.3109/09540261.2013.813442

Liu, J.-W., Tu, Y.-K., Lai, Y.-F., Lee, H.-C., Tsai, P.-S., Chen, T.-J., … Chiu, H.-Y. (2019). Associations between sleep disturbances and suicidal ideation, plans, and attempts in adolescents: a systematic review and meta-analysis. Sleep, 42(6). https://doi.org/10.1093/sleep/zsz054

Lopresti, A. L., Maker, G. L., Hood, S. D., & Drummond, P. D. (2014). A review of peripheral biomarkers in major depression: The potential of inflammatory and oxidative stress biomarkers. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 48, 102–111. https://doi.org/10.1016/j.pnpbp.2013.09.017

Malas, N., Plioplys, S., & Pao, M. (2019). Depression in Medically Ill Children and Adolescents. Child and Adolescent Psychiatric Clinics of North America, 28(3), 421–445. https://doi.org/10.1016/j.chc.2019.02.005

Martin, C., Tansey, K. E., Schalkwyk, L. C., & Powell, T. R. (2015). The inflammatory cytokines: Molecular biomarkers for major depressive disorder? Biomarkers in Medicine, 9(2), 169–180. https://doi.org/http://dx.doi.org/10.2217/BMM.14.29

Miller, B. R., & Hen, R. (2015). The current state of the neurogenic theory of depression and anxiety. Current Opinion in Neurobiology, 30, 51—58. https://doi.org/10.1016/j.conb.2014.08.012

Miller, D. B., & O'Callaghan, J. P. (2013). Personalized medicine in major depressive disorder - Opportunities and pitfalls. Metabolism: Clinical and Experimental, 62(SUPPL.1), S34–S39. https://doi.org/10.1016/j.metabol.2012.08.021

Monteggia, L. M., & Zarate, C. (2015). Antidepressant actions of ketamine: from molecular mechanisms to clinical practice. Current Opinion in

Neurobiology, 30, 139–143. https://doi.org/10.1016/j.conb.2014.12.004

Mullen, S. (2018). Major depressive disorder in children and adolescents. Mental Health Clinician, 8(6), 275–283. https://doi.org/10.9740/mhc.2018.11.275

Müller, N., Myint, A. M., & Schwarz, M. J. (2011). Inflammatory biomarkers and depression. Neurotoxicity Research, 19(2), 308–318. https://doi.org/10.1007/s12640-010-9210-2

Niciu, M. J., Ionescu, D. F., Richards, E. M., & Zarate, C. A. (2014). Glutamate and its receptors in the pathophysiology and treatment of major depressive disorder. Journal of Neural Transmission, 121(8), 907–924. https://doi.org/10.1007/s00702-013-1130-x

Okech, D., Hansen, N., Howard, W., Anarfi, J. K., & Burns, A. C. (n.d.). Social Support, Dysfunctional Coping, and Community Reintegration as Predictors of PTSD Among Human Trafficking Survivors. Behavioral Medicine (Washington, D.C.), 44(3), 209–218. https://doi.org/10.1080/08964289.2018.1432553

Orzechowska, A., Zajączkowska, M., Talarowska, M., & Gałecki, P. (2013). Depression and ways of coping with stress: a preliminary study. Medical Science Monitor : International Medical Journal of Experimental and Clinical Research, 19, 1050–1056. https://doi.org/10.12659/MSM.889778

Pedersen, G. A., Zajkowska, Z., Kieling, C., Gautam, K., Mondelli, V., Fisher, H. L., … Kohrt, B. A. (2019). Protocol for a systematic review of the development of depression among adolescents and young adults: psychological, biological, and contextual perspectives around the world. Systematic Reviews, 8(1), 179. https://doi.org/10.1186/s13643-019-1104-7

Pieters, T. (2017). The antidepressant era revisited; Towards differentiation and patient-empowerment in diagnosis and treatment. In G. Eghigian (Ed.), The Routledge History of Madness and Mental Health (2017th ed.). Retrieved from https://www.researchgate.net/publication/318852989_The_antidepressant_era_revisited_Towards_differentiation_and_patient-empowerment_in_diagnosis_and_treatment

Pieters, T., & Snelders, S. (2009, July 11). Psychotropic drug use: Between healing and enhancing the mind. Neuroethics, Vol. 2, pp. 63–73. https://doi.org/10.1007/s12152-009-9033-0

Raison, C. L., & Miller, A. H. (2011). Is depression an inflammatory disorder? Current Psychiatry Reports, 13(6), 467–475. https://doi.org/10.1007/s11920-011-0232-0

Roseman, L., Nutt, D. J., & Carhart-Harris, R. L. (2017). Quality of Acute Psychedelic Experience Predicts Therapeutic Efficacy of Psilocybin for Treatment-Resistant Depression. Frontiers in Pharmacology, 8, 974. https://doi.org/10.3389/fphar.2017.00974

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., … Fava, M. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. The American Journal of Psychiatry, 163(11), 1905–1917. https://doi.org/10.1176/ajp.2006.163.11.1905

Saxon, L., Makhashvili, N., Chikovani, I., Seguin, M., McKee, M., Patel, V., … Roberts, B. (2017). Coping strategies and mental health outcomes of conflict-affected persons in the Republic of Georgia. Epidemiology and Psychiatric Sciences, 26(3), 276–286. https://doi.org/10.1017/S2045796016000019

Schaakxs, R., Comijs, H. C., Lamers, F., Kok, R. M., Beekman, A. T. F., & Penninx, B. W. J. H. (2018). Associations between age and the course of major depressive disorder: a 2-year longitudinal cohort study. The Lancet Psychiatry. https://doi.org/10.1016/S2215-0366(18)30166-4

Sharpley, C. F., & Bitsika, V. (2013). Differences in neurobiological pathways of four "clinical content" subtypes of depression. Behavioural Brain Research, 256, 368–376. https://doi.org/10.1016/j.bbr.2013.08.030

Teismann, H., Wersching, H., Nagel, M., Arolt, V., Heindel, W., Baune, B. T., … Berger, K. (2014). Establishing the bidirectional relationship between depression and subclinical arteriosclerosis--rationale, design, and characteristics of the BiDirect Study. BMC Psychiatry, 14(1), 174. https://doi.org/10.1186/1471-244X-14-174

Uher, R., Dernovsek, M. Z., Mors, O., Hauser, J., Souery, D., Zobel, A., … Farmer, A. (2011). Melancholic, atypical and anxious depression subtypes and outcome of treatment with escitalopram and nortriptyline. Journal of Affective Disorders, 132(1–2), 112–120. https://doi.org/10.1016/j.jad.2011.02.014

Valkanova, V., Ebmeier, K. P., & Allan, C. L. (2013). CRP, IL-6 and depression: A systematic review and meta-analysis of longitudinal studies. Journal of Affective Disorders, 150(3), 736–744. https://doi.org/10.1016/j.jad.2013.06.004

Van Loo, H. M., Cai, T., Gruber, M. J., Li, J., De Jonge, P., Petukhova, M., … Kessler, R. C. (2014). Major depressive disorder subtypes to predict long-term course. Depression and Anxiety, 31(9), 765–777. https://doi.org/10.1002/da.22233

van Loo, H. M., de Jonge, P., Romeijn, J.-W., Kessler, R. C., & Schoevers, R. a. (2012). Data-driven subtypes of major depressive disorder: a systematic review. BMC Medicine, 10(1), 156. https://doi.org/10.1186/1741-7015-10-156

Vassilopoulou, K., Papathanasiou, M., Michopoulos, I., Boufidou, F., Oulis, P., Kelekis, N., … Lykouras, L. (2013). A magnetic resonance imaging study of hippocampal, amygdala and subgenual prefrontal cortex volumes in major depression subtypes: Melancholic versus psychotic depression.

Journal of Affective Disorders, 146(2), 197–204. https://doi.org/10.1016/j.jad.2012.09.003

Vibhakar, V., Allen, L. R., Gee, B., & Meiser-Stedman, R. (2019). A systematic review and meta-analysis on the prevalence of depression in children and adolescents after exposure to trauma. Journal of Affective Disorders, 255, 77–89. https://doi.org/10.1016/j.jad.2019.05.005

Wagner, M. T., Mithoefer, M. C., Mithoefer, A. T., MacAulay, R. K., Jerome, L., Yazar-Klosinski, B., & Doblin, R. (2017). Therapeutic effect of increased openness: Investigating mechanism of action in MDMA-assisted psychotherapy. Journal of Psychopharmacology (Oxford, England), 31(8), 967–974. https://doi.org/10.1177/0269881117711712

Watson, P. J., & Andrews, P. W. (2002). Toward a revised evolutionary adaptationist analysis of depression: the social navigation hypothesis. Journal of Affective Disorders, 72(1), 1–14. https://doi.org/10.1016/S0165-0327(01)00459-1

Weiss, J. M. (n.d.). Somatic effects of predictable and unpredictable shock. Psychosomatic Medicine, 32(4), 397–408.

Weiss, J. M. (1972). Psychological factors in stress and disease. Scientific American, 226(6), 104–113.

Weiss, J. M. (1991). Stress-induced depression: critical neurochemical and electrophysiological changes. In J. Madden (Ed.), Neurobiology of learning, emotion, and affect. (pp. 123–154). https://doi.org/10.1002/depr.3050010211

Willner, P., Muscat, R., & Papp, M. (1992). Chronic mild stress-induced anhedonia: a realistic animal model of depression. Neuroscience and Biobehavioral Reviews, 16(4), 525–534. https://doi.org/10.1176/appi.ajp.2011.11020335

World Health Organisation. (2016). Fact sheet Depression. Retrieved from Media centre website: http://www.who.int/mediacentre/factsheets/fs369/en/index.html

Yang, S. J., Stewart, R., Kang, H. J., Kim, S. Y., Bae, K. Y., Kim, J. M., … Jun, T. Y. (2013). Response to antidepressants in major depressive disorder with melancholic features: The CRESCEND study. Journal of Affective Disorders, 144(1–2), 42–50. https://doi.org/10.1016/j.jad.2012.06.004

Zimmerman, M., Walsh, E., Friedman, M., Boerescu, D. A., & Attiullah, N. (2017). Identifying Remission From Depression on 3 Self-Report Scales. The Journal of Clinical Psychiatry, 78(2), 177–183. https://doi.org/10.4088/JCP.16m10641

Zorn, J. V, Schür, R. R., Boks, M. P., Kahn, R. S., Joëls, M., & Vinkers, C. H. (2017). Cortisol stress reactivity across psychiatric disorders: A systematic review and meta-analysis. Psychoneuroendocrinology, 77, 25–36. https://doi.org/10.1016/j.psyneuen.2016.11.036

7.3 Evaluation

7.3.1 Strengths

The main strength of this study is that it starts with patients, and not with literature. This puts the patient central, and with it, the current clinical practice.

Literature can get lost in frameworks, hypothesis and theories that do not have a direct application for patients. Research in depression has tendencies to do that, as many frameworks do not translate to improvements for patients, evidenced by the fact that one third of all patients never reach a response after four lines of antidepressant treatment. There is a clear unmet need for some patients, so science needs to focus on why the existing treatments do not work and what to do to improve that.

7.3.2 Limitations

The main weakness is the lack of a systematic approach. It is a narrative review, without a formal search method. This means that other researchers cannot audit our selection, and it is likely that the selection is biased based on whether it agrees with our stance. We do not pretend that this is an unbiased overview, though, as that was not the goal. We aimed to shine a new light on a thoroughly researched field, to reach a breakthrough in thinking. Exclusively reviewing published research for that was not a good starting point.

To make sure that we presented our hypothesis properly, and compared it with existing theories, we included the views of several experts in their field and asked their opinion on our hypothesis, our presentation and the offered recommendations. Asking more experts for their opinion might have resulted in a more balanced presentation and set of recommendation, but might have reduced the bold ambition of this manuscript to change the treatment paradigm.

7.3.3 Appropriateness of the methodology

Before we started this narrative review, we considered a systematic review. It seemed impossible to achieve a breakthrough insight using existing information, so we decided to change the scope. Full network analyses on antidepressants had been performed, but had not yielded insights that significantly helped the research on treating depression forward. Therefore, we decided to use a different approach, to generate

a new hypothesis and put the patient central. We disseminated two patient cases to consider how patients can be treated better and added literature where relevant.

Using this approach allowed us to choose what we would include without constraints. So next to the patient cases, we included literature from several scientific disciplines, such as evolutionary biology, clinical epidemiology and neurochemistry. The result is a narrative review, that does not pretend to be a comprehensive review of the literature. Instead, it offers a new way of thinking about adolescent depression. Whether this is beneficial for patients would need to be tested in further studies, so this is not a definitive answer to the research question.

7.3.4 Lessons learned

This review showed that an SR is not the most appropriate research method for all questions. We require new hypotheses to generate a breakthrough, so a narrative review was more appropriate.

Future research could dive deeper into coping mechanisms and consider how to improve them for patients who might benefit from that. It could also look at how to identify patients at an earlier stage than during clinical depressions, as prevention might be much more effective than treatment. Evidence needs to be generated to support or falsify this hypothesis first before new SRs and meta-analytic techniques can be used.

8 Discussion

In this thesis, we show examples of how to tailor the methodology of SRs to find the relevant sources and achieve a meaningful answer to each research question. This methodology is important, whether the goal is a preliminary overview of the literature in a scoping review, a full systematic review, a meta-analysis, or even a full meta-regression and network analyses. Aspects to consider are which databases are used for the search, how search queries are formulated, which in- and exclusion criteria are applied, and how the resulting data are synthesised and presented.

In chapter 3, 4 and 5 we looked at which variations in search methodology are possible. Specifically, in chapter 3 we showed that using three databases did not duplicate our selection, as only 299 out of 3,508 records – or 8.5% – were duplicates. Each database added unique material that would not have been found without searching that database. Having two reviewers select relevant papers and discuss their differences helped in a robust selection, so that fewer relevant papers were missed.

We also showed that a completely tailored search string led to a higher precision of the search than a short, general search string, though still only 126 – or 3.6% – of records were deemed relevant. In comparison, chapter 4 included 172 out of 2,115 or 8.1% of found records, chapter 5 used 161 out of 11,059 or 1.5%, and chapter 6 included 114 out of 754 or 15.1% of records. This shows that the fraction of included research varies heavily between research areas. Notable is also that the question pertaining to the most well-indexed research field, on dietary fibres, produced the most accurate search results.

Chapter 4 showed that searching non-scientific databases can yield other perspectives than the scientific literature only can generate, leading to a more balanced overall judgement. Which newspapers are searched, and how these searches are performed, will require attention in the design of an SR as it impacts the perspective and results.

In chapter 5 we took another approach, adjusting the search methodology for each protein. Though this meant we had a pre-selected list of proteins, which may not have been complete, this allowed us to find the most relevant material for each. If we had first performed a search to find more likely targets, we might have found targets that we

now missed out on. But as the goal was to assess the prediction that gene doping with these targets would have become more common in the last five years, this was an appropriate method for this question.

Furthermore, in chapter 5, we looked at how to integrate qualitative data. We developed a scoring method to estimate which proteins are most likely to be targeted for gene doping first. We can further refine this method by having independent reviewers score all proteins first, and then reconcile the differences. This is likely to be more reliable than the current method of drafting and reviewing, which may lead to anchoring bias. We could also improve this by clearly defining the criteria and definitions for the categories that are scored on.

None of these SRs looked at evidence selection bias, or specifically publication bias. Because the data were not comparable enough, it was impossible to quantitatively assess the possible existence of more data that had not been published. In chapter 6, where we used quantitative data, we did assess publication bias with a funnel plot.

We looked at the impact of dietary fibres on cholesterol in animal research, using the same modified search strategy as chapter 5. This dataset allowed us to collect quantitative data and estimate an overall effect. To assess publication bias, we presented a funnel plot. This allowed us to assess whether we had a full picture of the research, or if there was a selection in what had been published. This exposed likely publication bias, indicating that the effects we found were likely an overestimate of the real effect.

These SRs were complemented with a narrative review in chapter 7, to highlight that these have value too for answering select questions. We also wrote letters to editors which underlines our expertise in the field we previously published on.

Not all methods we chose resulted in the most efficient or thorough way to investigate a specific question. For all questions, more research on the topic in advance of selecting criteria and articles would have been beneficial, and the consistent availability of a duplicated selection and scoring by an independent author would have been helpful.

The best way to assess the impact of the method on the outcomes of an SR would be to perform an SR on a question in several different methods, and then compare the results. This would take significant time and would require independent teams of researchers, making it not

feasible for this thesis. Another option is to look at a situation in which this has happened organically: research questions that have been covered by multiple independent SR researchers. In paragraph 2.3.2, we discuss a clear example of this: there are many SRs about the impact of alcohol on mortality. In this example, the inclusion of ex-alcoholics as current abstainers consistently skewed the findings towards a J-curve, with optimal outcomes for those drinking low amounts of alcohol. This was not replicated in analyses where ex-alcoholics were excluded. This real-world scenario highlights what we hypothesised: the method of SRs has an impact on the outcomes and needs to be carefully designed.

Different authors might make different choices, which explains why different SRs on the same topic sometimes find different or outright conflicting outcomes. Just like for clinical trials, it is necessary to have a debate about and a rationale for why specific criteria are chosen. This should encompass all the design features that are laid out in this thesis, and will increases the likelihood that SRs find the relevant effect.

For policy makers, SRs can be a useful tool to summarize and compare findings of a policy proposal in different settings. Different datasets will be needed, and expertise to scrutinize the research methods and the implemented policies, to reliably summarize and consider the value of the literature. The comparability of the policy methods needs to be assessed thoroughly though, as the fine details in seemingly comparable policy can make a difference. Similarly, the details of the setting in which a policy has been implemented needs to be considered. But finding all available research on a specific policy proposal, and comparing outcomes, is a valuable method to consider while designing a new policy measure, before implementing.

In time, with the maturing of the data and tools to compare heterogeneous selections of studies, it might be possible to perform network meta-analyses to indirectly compare health-related research questions that are not as easily captured as those in EBM, and even perform meta-regressions on some. This will require more standardisation in research areas, and clear definitions of concepts, so for now scoping reviews will be more common.

In the following parts of the discussion, we will debate some remaining questions, and consider some possible answers. First, we look into several threats that could compromise the credibility of SRs,

whether in EBM or in health-related research outside EBM. Then we discuss the fundamental credibility of EBM itself. We look at real-world evidence, which faces very similar challenges as SRs do, and at the future of SRs as artificial intelligence and further computerization progress into the field of healthcare. Finally, we consider the relevance of the intertwined history of EBM and SR to the debate today.

8.1 Translation of challenges: real-world evidence

The quickly developing field of real-world evidence (RWE) faces very similar challenges to SRs. RWE is the use of data from real-world sources, such as electronic health records, hospital files or insurance claims, to find the effects of a treatment on patients in the real world.[34] It answers the criticism that clinical trials are too "perfect," with a stringent set of inclusion criteria that do not reflect the real-world population. This leaves an evidence gap for patients with different characteristics than those included in a trial, with, for example, comorbidities, polypharmacy, different ages or organ disfunctions.

Also, RWE helps to see if the implementation of a drug in a real-world setting matches up with the trial data. This type of evidence is increasingly being considered more valuable, and regulators are starting to accept data from RWE studies on off-label use to consider new indications of an existing drug instead of requesting formal trials. This speeds up evidence generation, as these studies are much faster than clinical trials, and it might make development of drugs more affordable. The ISPE-ISPOR taskforce on RWE in regulatory decision-making works on recommendations to make RWE more useful and reliable for regulatory authorities.[180]

Though the use and acceptance of RWE is a valuable addition to the arsenal of tools that are available to estimate a treatment effect, there are similar problems as with SRs. Right now, the largest datasets are being bought, cleaned and analysed by a small set of companies, who sell these datasets to pharmaceutical companies and other parties with an interest. These can then, selectively, choose to analyse them as they see fit, and publish the ones that support their messaging. This leads to a misrepresentation of the research field in the favour of the owners, which makes the field less reliable.

A solution to this would be to have an independent research organisation track the outcomes of drugs in the real world, so that the

results are published regardless of whether the outcomes are favourable. With improved computerization of healthcare, and increased data on real-world treatment patterns, this seems increasingly possible. The development of this organisation would be in line with the Cochrane collaboration for SRs, and could demand the publication of protocols before analyses on a forum similar to EudraCT, clinicaltrials.gov or PROSPERO, which would allow for public scrutiny and oblige the organisation to publish the results. With initiatives such as ENCePP and their EU PAS register for non-interventional research this is being implemented.

Even then, it is possible to only register once the data is analysed and a positive outcome is expected, similar to PROSPERO for SRs. The most reliable option would still be to have independent researchers perform studies.

8.2 Threats to credibility of SRs

There are some threats to the credibility of SRs, though. In recent years, SRs have become abundant[15,81,108,181] with many stakeholders involved in the development. From altruistic reasons, to scholarly credits and academic appreciation, to policy advisers and business professionals, to outright marketing professionals, all kinds of different people can perform SRs.[81] These stakeholders have different incentives, though, so judging the appropriateness of the methodology of each SR to understand the reliability of the results is essential.

8.2.1 Authorship for sale

One threat to the credibility of SRs is the recent flood of SRs with MA, mostly from China,[81,182] with close to identical study setup and phrasing ("…Such studies taking these factors into account may eventually lead to our better, comprehensive understanding…"). Also, some very specific errors appear ("Begger's funnel plot", a combination of Colin Begg and Matthias Egger's names, both statisticians who worked on publication bias).[183] This indicates that meta-analyses are carried out as a routine exercise, with changes in genes, indication and mostly authors, as a paid service to investigators looking for prestigious authorships. Though they seem credible, the results of these low-quality SRs are often dubious due to methodological shortcomings. Journal editors should use plagiarism detection tools to detect these types of papers, but that will not solve this problem completely. To limit this, editors and peer reviewers assess

papers more carefully before publishing, flag suspicious late changes in authorship and request more data from the authors.[183]

8.2.2 Industry-funded SRs

Similarly, industry-funded SRs and MAs are increasing in abundance. They are more likely to have a positive conclusion and leave out side-effects,[130] skewing the landscape of scientific literature to more favourable opinions. This shows that SRs, just like clinical trials and other publications, can be misused as a marketing tool, and should be scrutinised for quality like other studies.[130]

8.2.3 Pace

Another threat is the high pace of scientific progress. New drugs and treatments are introduced at a high pace, which can make an SR outdated soon after publishing. Though there will always be a delay between the dissemination of individual studies and their aggregation, due to the laborious tasks associated with performing an SR, this threatens the reliability of the findings.

8.2.4 Quality

Finally, a major threat for SRs is that, when unreliable methodologies are used to aggregate data, and the results are not reliable, the entire body of literature focussed on compiling research might be perceived as less reliable.[184] This would pander to those who saw SRs with MA as 'statistical trickery' during the introduction of the method. Given that low-quality SRs are more likely to find a significant effect, comparable to positive results in RCTs,[115,141,184] and positive results might be more likely to be published, there might be a publication bias favouring low-quality SRs. This is possibly due to the influence of sponsors or the interest editors and peer-reviewers have in publishing positive data, similar to the causes of publication bias in clinical trials. This bias damages the reputation of the field and undermines the implementation of EBM.

8.3 Solutions to those threats

8.3.1 Improve quality by using checklists

Editors should require a PRISMA checklist,[157] and use or ask peer-reviewers to use the Oxman and Guyatt scoring checklist[108,185] or a similar tool for estimating quality.[184] This will help identify reliable SRs and MAs, and help improve submitted manuscripts.

8.3.2 Reduce publication bias

A critical tool in the reduction of publication bias for primary research was the introduction of clinical trial registries. This meant that if a clinical trial is started, there is an obligation to report the results. The introduction of PROSPERO was initially hailed as a similar move, with a similar impact.

The results of the introduction of PROSPERO have been somewhat less impactful though. Currently, only a small fraction (possibly 10%) of published SRs are registered with PROSPERO.[81] This is partially due to the restrictions of PROSPERO, as it does not accept registration of systematic reviews that do not clearly relate to human health. Several of our reviews would not have been accepted for PROSPERO registration. PROSPERO could broaden its remit to include more SRs.

Furthermore, given that the information needed to perform an SR is publicly available, it is perfectly possible to perform a draft SR, and if the outcomes are favourable, register, perform a full SR, and publish.[81,118] And as many SR researchers are independent, unlike with clinical trials, forcing the publication of the results might not improve this. PROSPERO could offer to host the results of SRs that are rejected by journals, as do clinical trial registries, to make sure journal rejections are not a reason for lack of disclosure.

The same problem exists with preclinical research and real-world data, though with results that do not affect patients as directly. The only method of dealing with this is estimating publication bias retrospectively and considering the impact.

To have independent data available, research organisations and guideline committees are performing their own SRs and MAs. This helps answer clinical questions and translates literature to digestible clinical recommendations.

8.3.3 Increase independent funding

MAs with at least one industry-employed author are 22 times more likely not to include a negative statement about the assessed drug in their conclusion.[130] Part of the success of the Cochrane collaboration is due to the fact that the investigators receive no financial compensation, making them and their work credible. Improved funding options would also improve the number of recent SRs, which would make more up-to-date

SRs available to guide a decision. This will in turn help to more frequently update clinical guidelines, many of which are currently updated only every five years.

Similarly, more clinical trials without industry funding are needed. This would help address research gaps that have no patentable outcome (such as the effect of exercise for depression[186]), and add an independent voice to the literature. Similar to how politicians and athletes are not expected to provide their own scores, companies should not be asked to provide data that they have an interest in.[8]

A global registry for independent SRs would also help address this. That would make it easier to find SRs that have not been funded by stakeholders such as the pharmaceutical and medical device industry, increasing the ease of finding credible SRs for HCPs.

8.4 Future of SRs and EBM

As SRs are labour-intensive to produce, there will be a move to have more of the work automated. There has been progress in using more specific tools to aid with search, selection, synthesis and citation, but as more tools become available, a more comprehensive model can be imagined by their integration.[187] Already, reference managers that help keep track of search results, selection, and even obtaining full texts are available, and data integration is aided with specific analytical software. Future developments could be more automated. For example, natural language processing, a field of research in computer science, might help with the analysis of written text. Natural language generation is a technique that might even perform some of the writing, similar to how some news articles are already written. This will help with one of the main weaknesses of SRs: the high need for resources to keep up with the pace of evidence generation.

The writing of systematic reviews might become more automated, similar to how IBM's supercomputer Watson integrates all the available literature and treatment histories of similar patients to help with the selection of a treatment plan for an individual cancer patient.[188] In the long term, these systems might even take over a significant part of the tasks of healthcare providers.

One of the early criticisms of meta-analyses remains important to keep in mind: that effects can be obscured or made up by statistical alchemy. When compiling information for different studies, the scientific

rationale for comparing them needs to be clear, especially when using more advanced methods such as network analyses or meta-regressions. Definitions and concepts need to be clear beyond possible dispute, so scoping reviews are necessary to iron those out where they are not clear yet.

For now, though, caution is needed, as it is difficult to train a system as comprehensively as human HCPs or SR writers have been trained, so errors can be significant. Watson recently came in the news for selecting unsafe and inappropriate treatment plans.[189] An appropriately trained human will remain essential for the design of SRs and the implementation of their findings, to translate data from efficacy to efficiency to effectiveness.

9 Conclusion

Scientific progress is based on 'standing on the shoulders of your predecessors', and as a historic edifice must be based on cumulative efforts.[190] It is important to have a complete overview of the accomplishments of the work of predecessors, for which SR have been deliberately designed.[79]

The use of SRs and the development of their formal methodology has made EBM possible. With an ever-growing number of publications, SRs are a seminal tool for EBM. SRs have improved clinical practice substantially, but the use of SRs has not consistently spread to other research areas that impact health. Applying the methodology of SRs to health-related disciplines outside EBM makes it possible to improve research in these areas as well. This requires ensuring that the methods are designed as appropriate for the research question.

For readers of health-related SRs, it is important to consider the methods of the paper. Though reviews can be a quick way to read about all the science to answer a question, it is important to consider how reliable the overview is. Especially for policy makers or decision makers, biased information can impact significant groups of people.

As this thesis demonstrates, SRs outside EBM are valuable and feasible, but the methodology needs to be tailored at several stages. In each chapter we looked at options to make the method fit the question best, which provided key lessons about how to optimize SR methodology.

First, determining in which databases to search is relevant. A broader range of options must be considered, as not all data might be captured in the usual scientific datasets. For example, in chapter 3, we included PsychInfo, a specialised search engine for behavioural science and mental health, next to the broader search engines PubMed and Embase. In chapter 4, we included newspaper articles, to make sure we captured all relevant material.

Second, the search terms need to be tailored, as thesauri are not always suited for non-clinical questions. In chapter 3, we considered using the thesauri of three search engines and tailoring it by selecting all seemingly relevant terms. This improved the selection, so for areas which are not indexed as well as core clinical research areas this might be a viable option. For other questions, such as the one in chapter 4using

a short query that captures all relevant material and selecting from that is the best option. Using a short string with one interchangeable term proved effective in chapters 5 and 6.

Third, the means of integrating data varies, as the tools used in EBM are not always applicable. Compiling data with forest plots and summary statistics is not always possible, let alone more advanced methods as meta-regression or network analyses, so an appropriate method for scoring needs to be found. In chapter 5, we designed a scoring mechanism, and in chapter 6, we compiled the available data similar to how it would be done in EBM. Many questions will call for a scoping review or an SR without quantitative integration. Specifically for policy-related questions, integration of the data is not likely to feasible given the complexity of the policy itself and the lack of comparability of the landscape it is implemented in. A more descriptive approach will likely remain the most appropriate.

Fourth, estimating publication bias can be difficult. As there are no compulsory registries for trials outside EBM, there is no visibility of what has not been published. New registries for observational research are filling this void but are not mature enough to reliably establish what has been researched. Furthermore, quantitative tools to assess publication bias cannot be used for non-quantifiable data, which is common in research areas like ours, outside EBM. In chapter 6, we compiled the data from trials with different animal species, fibres and methods to look at this bias, and despite the shortcomings of the data, we show that publication bias likely caused an overestimation of the effects.

Finally, we also found a scenario in which an SR was not appropriate at all. Looking for a new hypothesis in chapter 7, we did not manage to find a new insight through our SR methodology. So, we changed the approach altogether, and started with patient cases and a select body of literature. This did not yield a straightforward evidence-based conclusion, but it did offer new insights.

As these studies highlight, using an appropriate methodology for SRs is important, similar to the importance of designing the right methodology for a clinical trial. A good design is more likely to result in unbiased, evidence-based outcomes. Even so, the increase of published systematic reviews over the last decades has not led to a consistent

increase in quality,[15,108,115,139,141,181,184] so improvement of their methodology is still warranted.

The fundamental question, on what the right balance is between aggregate data versus individual observations, is still relevant. Each treatment or lifestyle decision for an individual patient is a trial; the place of biomedical science is to make sure that the option that has the highest chance of succeeding gets the first chance, so that the benefit-risk ratio is optimized. SRs help make knowledge about the best choice available.

Our hypothesis, that each research question requires a tailored methodological approach to find reliable results, appears to be valid. This thesis shows examples of how to tailor the methodology for five research questions, and details key design features that can impact which data is found and how the results are compiled. It is a step in the direction of developing methods for SRs for health-related research questions EBM.

In conclusion, the methodology of SRs for health-related questions requires tailoring for the most reliable outcomes. It will remain necessary to evaluate and debate the appropriateness of a methodology for each research question, which requires expertise in SR methodology. As SRs become more commonly used for all health-related research questions, and meta-analytic techniques will become more possible, they will improve the availability of scientific findings to HCPs and the general population, and in term will lead to better health outcomes.

10 References

1. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* **268**, 2420–5 (1992).
2. Levin, A. The Cochrane Collaboration. *Ann. Intern. Med.* **135**, 309–12 (2001).
3. Skelly, A. C. & Chapman, J. Evidence-based medicine (EBM): origins and modern application to spine care. *Evid. Based. Spine. Care. J.* **2**, 11–6 (2011).
4. Harris, J. D., Quatman, C. E., Manring, M. M., Siston, R. A. & Flanigan, D. C. How to Write a Systematic Review. *Am. J. Sports Med.* **42**, 2761–2768 (2014).
5. Torpy, J. M. Evidence-Based Medicine. *JAMA J. Am. Med. Assoc.* **296**, 1192–1192 (2006).
6. Gopalakrishnan, S. & Ganeshkumar, P. Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. *J. Fam. Med. Prim. care* **2**, 9–14 (2013).
7. Dickersin, K., Straus, S. E. & Bero, L. A. Evidence based medicine: increasing, not dictating, choice. *BMJ* **334**, s10–s10 (2007).
8. Every-Palmer, S. & Howick, J. How evidence-based medicine is failing due to biased trials and selective publication. *J. Eval. Clin. Pract.* **20**, 908–14 (2014).
9. van der Geest, M. Sjoerd Repping, de man die alle bewezen onzinzorg uit de ziekenhuizen moet schrappen. *De Volkskrant* (2019).
10. Crowther, M. & Lim, W. Systematic review and meta-analysis methodology. *Blood* **116**, 3140–3146 (2010).
11. The Cochrane Collaboration. Systematic reviews. in *Cochrane Handbook for Systematic Reviews of Interventions* (eds. Higgins, J. & Green, S.) (Cochrane Community, 2011).
12. Fletcher, R. H. & Fletcher, S. W. Evidence-based approach to the medical literature. *J. Gen. Intern. Med.* **12 Suppl 2**, S5-14 (1997).
13. BMJ. Medical milestones: Celebrating key advances since 1840. *Br. Med. J.* **334**, S1–S22 (2007).
14. Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B. & Richardson, W. S. Evidence based medicine: what it is and what it isn't. *BMJ* **312**, 71–2 (1996).
15. Jadad, A. R. *et al.* Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. *BMJ* **320**, 537–40 (2000).
16. Sehon, S. R. & Stanley, D. E. A philosophical analysis of the evidence-based medicine debate. *BMC Health Serv. Res.* **3**, 14 (2003).
17. Rangachari, P. K. Evidence-based medicine: old French wine with a new Canadian label? *J. R. Soc. Med.* **90**, 280–4 (1997).
18. Accad, M. & Francis, D. Does evidence based medicine adversely affect clinical judgment? *BMJ* k2799 (2018). doi:10.1136/bmj.k2799

19. Ruschitzka, F. Cardiac Resynchronisation Therapy. *EUROPEAN HOSPITAL* **17**, 613–621 (2007).
20. Ruschitzka, F. *et al.* Cardiac-resynchronization therapy in heart failure with a narrow QRS complex. *N. Engl. J. Med.* **369**, 1395–405 (2013).
21. The Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions*. (2011).
22. Herrett, E. *et al.* Study protocol for statin web-based investigation of side effects (StatinWISE): a series of randomised controlled N-of-1 trials comparing atorvastatin and placebo in UK primary care. *BMJ Open* **7**, e016604 (2017).
23. Stolberg, H. O., Norman, G. & Trop, I. Randomized controlled trials. *AJR. Am. J. Roentgenol.* **183**, 1539–44 (2004).
24. Berlin, J. A. & Golub, R. M. Meta-analysis as evidence: building a better pyramid. *JAMA* **312**, 603–5 (2014).
25. Dechartres, A., Altman, D. G., Trinquart, L., Boutron, I. & Ravaud, P. Association Between Analytic Strategy and Estimates of Treatment Outcomes in Meta-analyses. *JAMA* **312**, 623 (2014).
26. Howick, J. *et al.* Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document). *Oxford Centre for Evidence-Based Medicine* Available at: https://www.cebm.net/index.aspx?o=5653. (Accessed: 9th January 2019)
27. De Brún, C. *Finding the Evidence: A key step in the information production process*. *The Information Standard Guide* (2013).
28. Smith, G. C. S. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *Bmj* **327**, 1459–1461 (2003).
29. Benson, K. & Hartz, A. J. A comparison of observational studies and randomized, controlled trials. *N. Engl. J. Med.* **342**, 1878–86 (2000).
30. Shah, H. M. & Chung, K. C. Archie Cochrane and his vision for evidence-based medicine. *Plast. Reconstr. Surg.* **124**, 982–8 (2009).
31. Cochrane, A. *Effectiveness and efficiency: Random reflections on health services*. (Nuffield Trust, 1972).
32. Yeh, R. W. *et al.* Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. *BMJ* **363**, k5094 (2018).
33. Hoyme, U. B. Pragmatic prevention of preterm birth and evidence based medicine. *Arch. Gynecol. Obstet.* **294**, 1–3 (2016).
34. Staa, T.-P. v. *et al.* Pragmatic randomised trials using routine electronic health records: putting them to the test. *BMJ* **344**, e55–e55 (2012).
35. Vandenbroucke, J. P. Why do the results of randomised and observational studies differ? *BMJ* **343**, d7020–d7020 (2011).
36. Vandenbroucke, J. P. Clinical epidemiology: A daydream? *Eur. J. Epidemiol.* **32**, 95–101 (2017).
37. MD|OG. Mijn Data Onze Gezondheid. (2019). Available at:

https://mdog.nl/.
38. Vandenbroucke, J. P. When are observational studies as credible as randomised trials? *Lancet* **363**, 1728–1731 (2004).
39. Murad, M. H., Asi, N., Alsawas, M. & Alahdab, F. New evidence pyramid. *Evid. Based Med.* **21**, 125–127 (2016).
40. Tugwell, P. & Knottnerus, J. A. Is the 'Evidence-Pyramid' now dead? *J. Clin. Epidemiol.* **68**, 1247–50 (2015).
41. Baker, K. A. & Weeks, S. M. An Overview of Systematic Review. *J. PeriAnesthesia Nurs.* **29**, 454–458 (2014).
42. RELX. Systematic review. *EMBASE EMTREE* (2004). Available at: https://www.embase.com/#emtreeSearch/search/72829::systematic review. (Accessed: 19th August 2018)
43. National Center for Biotechnology Information - U.S. National Library of Medicine. Review [Publication Type]. *MESH* (2008). Available at: https://www.ncbi.nlm.nih.gov/mesh/68016454. (Accessed: 19th August 2018)
44. Yuan, Y. & Hunt, R. H. Systematic reviews: the good, the bad, and the ugly. *Am. J. Gastroenterol.* **104**, 1086–92 (2009).
45. Chalmers, I., Hedges, L. V & Cooper, H. A brief history of research synthesis. *Eval. Health Prof.* **25**, 12–37 (2002).
46. Grant, M. J. & Booth, A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info. Libr. J.* **26**, 91–108 (2009).
47. Petkovic, J., Welch, V. & Tugwell, P. Do evidence summaries increase policy-makers' use of evidence from systematic reviews: A systematic review protocol. *Syst. Rev.* **4**, 2–7 (2015).
48. Warner, J. H. & American Council of Learned, S. *The therapeutic perspective : medical practice, knowledge, and identity in America, 1820-1885*. (1997).
49. Semmelweis, I. *The etiology, concept, and prophylaxis of childbed fever.* (The University of Wisconsin Press, 1983).
50. Evidence-based medicine, in its place. *Lancet (London, England)* **346**, 785 (1995).
51. Greenhalgh, T., Howick, J., Maskrey, N. & Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ* **348**, g3725 (2014).
52. Feinstein, A. R. 'Clinical Judgment' revisited: the distraction of quantitative models. *Ann. Intern. Med.* **120**, 799–805 (1994).
53. Manser, R. & Walters, E. H. What is evidence-based medicine and the role of the systematic review: the revolution coming your way. *Monaldi Arch. chest Dis. = Arch. Monaldi per le Mal. del torace* **56**, 33–8 (2001).
54. Collier, R. Legumes, lemons and streptomycin: a short history of the clinical trial. *CMAJ* **180**, 23–4 (2009).
55. Lind, J. *A treatise of the scurvy. In three parts. Containing an inquiry into the nature, causes and cure, of that disease. Together with a critical and*

chronological view of what has been published on the subject. (Sands, Murray and Cochran, 1753).
56. Pieters, T. & Widdershoven, G. *Filosofie & geschiedenis van de gezondheidszorg*. (Boom uitgevers, 2019).
57. Louis, P. & Jackson, J. *Researches on the effects of bloodletting in some inflammatory diseases, and on the influence of tartarized antimony and vesication in pneumonitis*. (Boston [Mass.] : Hilliard, Gray, 1836).
58. Best, M. Pierre Charles Alexandre Louis: Master of the spirit of mathematical clinical science. *Qual. Saf. Heal. Care* **14**, 462–464 (2005).
59. Louis, P. Researches on the Effects of Bloodletting in Some Inflammatory Diseases, and on the Influence of Tartarized Antimony and Vesication in Pneumonitis. *Br. foreign Med. Rev.* **3**, 456–459 (1837).
60. Ierodiakonou, K. & Vandenbroucke, J. P. Medicine as a stochastic art. *Lancet (London, England)* **341**, 542–3 (1993).
61. Fernández-Guerrero, I. M., Torralbo, M. & Fernández-Cano, A. A Forerunner of Qualitative Health Research. *Qual. Health Res.* **24**, 124–135 (2014).
62. Bernard, C. *Introduction à l'étude de la médecine expérimentale*. (JB Bailliere et fils, 1865).
63. Bernard, C. *An introduction to the study of experimental medicine / Translated by H.C. Greene*. (Macmillan & Co, 1927).
64. Murphy, T. D. Medical knowledge and statistical methods in early nineteenth-century France. *Med. Hist.* **25**, 301–19 (1981).
65. Morabia, A. Claude Bernard, statistics and comparative trials. *J. R. Soc. Med.* **111**, 335–336 (2018).
66. DePalma, R. G., Hayes, V. W. & Zacharski, L. R. Bloodletting: Past and Present. *J. Am. Coll. Surg.* **205**, 132–144 (2007).
67. Pearson, K. Report on Certain Enteric Fever Inoculation Statistics. *Br. Med. J.* **2**, 1243–6 (1904).
68. Olkin, I. Meta-analysis: current issues in research synthesis. *Stat. Med.* **15**, 1253-7; discussion 1259-62 (1996).
69. Edwards, A. W. F. *Statistical methods for research workers*. (Oliver and Boyd Ltd., 1925).
70. Hill, G. B. Archie Cochrane and his legacy. An internal challenge to physicians' autonomy? *J. Clin. Epidemiol.* **53**, 1189–92 (2000).
71. Cochrane, A. Sickness in Salonica: my first, worst, and most successful clinical trial. *Br. Med. J. (Clin. Res. Ed).* **289**, 1726–7 (1984).
72. Chalmers, I. Unbiased, relevant, and reliable assessments in health care: important progress during the past century, but plenty of scope for doing better. *BMJ* **317**, 1167–8 (1998).
73. Crofton, J. The MRC randomized trial of streptomycin and its legacy: a view from the clinical front line. *J. R. Soc. Med.* **99**, 531–4 (2006).
74. Medical Research Council. STREPTOMYCIN treatment of tuberculous meningitis. *Lancet* **1**, 582–96 (1948).
75. Yoshioka, A. Use of randomisation in the Medical Research Council's

clinical trial of streptomycin in pulmonary tuberculosis in the 1940s. *BMJ* **317**, 1220–3 (1998).
76. HILL, A. B. The clinical trial. *N. Engl. J. Med.* **247**, 113–9 (1952).
77. Munn, Z. *et al.* Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med. Res. Methodol.* **18**, 143 (2018).
78. DANIELS, M. & HILL, A. B. Chemotherapy of pulmonary tuberculosis in young adults; an analysis of the combined results of three Medical Research Council trials. *Br. Med. J.* **1**, 1162–8 (1952).
79. Feldman, K. A. Using the Work of Others: Some Observations on Reviewing and Integrating. *Sociol. Educ.* **44**, 86–102 (1971).
80. Mulrow, C. D. The medical review article: state of the science. *Ann. Intern. Med.* **106**, 485–8 (1987).
81. Ioannidis, J. P. A. A. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q.* **94**, 485–514 (2016).
82. TAYLOR, B. N., PARKER, W. H. & LANGENBERG, D. N. Determination of e/h, Using Macroscopic Quantum Phase Coherence in Superconductors: Implications for Quantum Electrodynamics and the Fundamental Physical Constants. *Rev. Mod. Phys.* **41**, 375–496 (1969).
83. Antiplatelet Trialists' Collaboration. Secondary prevention of vascular disease by prolonged antiplatelet treatment. Antiplatelet Trialists' Collaboration. *Br. Med. J. (Clin. Res. Ed).* **296**, 320–31 (1988).
84. Wachter, K. W. Disturbed by meta-analysis? *Science* **241**, 1407–8 (1988).
85. Eysenck, H. J. An exercise in mega-silliness. *Am. Psychol.* **33**, 517–517 (1978).
86. Thompson, S. G. & Pocock, S. J. Can meta-analyses be trusted? *Lancet (London, England)* **338**, 1127–30 (1991).
87. Shapiro, S. Meta-analysis/Shmeta-analysis. *Am. J. Epidemiol.* **140**, 771–8 (1994).
88. Feinstein, A. R. Meta-analysis: statistical alchemy for the 21st century. *J. Clin. Epidemiol.* **48**, 71–9 (1995).
89. Light, R. J. & Pillemer, D. B. *Summing Up. The Science of Reviewing Research*. (Harvard University Press, 1984).
90. Misra, D. P. & Agarwal, V. Systematic Reviews: Challenges for Their Justification, Related Comprehensive Searches, and Implications. *J. Korean Med. Sci.* **33**, e92 (2018).
91. Egger, M., Davey Smith, G., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* **315**, 629–34 (1997).
92. Clarke, M., Brice, A. & Chalmers, I. Accumulating Research: A Systematic Account of How Cumulative Meta-Analyses Would Have Provided Knowledge, Improved Health, Reduced Harm and Saved Resources. *PLoS One* **9**, e102670 (2014).
93. Leimu, R. & Koricheva, J. Cumulative meta-analysis: a new tool for detection of temporal trends and publication bias in ecology.

Proceedings. Biol. Sci. **271**, 1961–6 (2004).
94. Lau, J. *et al.* Cumulative Meta-Analysis of Therapeutic Trials for Myocardial Infarction. *N. Engl. J. Med.* **327**, 248–254 (1992).
95. Cochrane, A. 1931-1971: a critical review with particular reference to the medical profession. in *Medicines for the year 2000* 1–11 (Office of Health Economics, 1979).
96. Sawers, N. Evidence-Based Medicine vs Traditional Healers in Africa. *JAMA Ophthalmol.* **134**, 1085–1086 (2016).
97. Stanley, C. C. *et al.* Risk factors and reasons for treatment abandonment among children with lymphoma in Malawi. *Support. Care Cancer* **26**, 967–973 (2018).
98. Ramirez, P. T. *et al.* Minimally Invasive versus Abdominal Radical Hysterectomy for Cervical Cancer. *N. Engl. J. Med.* **379**, 1895–1904 (2018).
99. Melamed, A. *et al.* Survival after Minimally Invasive Radical Hysterectomy for Early-Stage Cervical Cancer. *N. Engl. J. Med.* **379**, 1905–1914 (2018).
100. Jonas, W. B. *et al.* To what extent are surgery and invasive procedures effective beyond a placebo response? A systematic review with meta-analysis of randomised, sham controlled trials. *BMJ Open* **5**, e009655 (2015).
101. Rother, E. T. Systematic literature review X narrative review. *Acta Paul. Enferm.* **20**, v–vi (2007).
102. Cook, D. J., Mulrow, C. D. & Haynes, R. B. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann. Intern. Med.* **126**, 376–80 (1997).
103. Greenhalgh, T. Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ* **315**, 672–5 (1997).
104. Pauling, L. *How to Live Longer and Feel Better*. (W.H. Freeman and Company, 1986).
105. Knipschild, P. Systematic reviews. Some examples. *BMJ* **309**, 719–21 (1994).
106. Zhen, J. T. *et al.* Genetic testing for hereditary prostate cancer: Current status and limitations. *Cancer* 3105–3117 (2018). doi:10.1002/cncr.31316
107. Paul, M. & Leibovici, L. Systematic review or meta-analysis? Their place in the evidence hierarchy. *Clin. Microbiol. Infect.* **20**, 97–100 (2014).
108. Dijkman, B. G. *et al.* Twenty years of meta-analyses in orthopaedic surgery: has quality kept up with quantity? *J. Bone Joint Surg. Am.* **92**, 48–57 (2010).
109. Colquhoun, H. L. *et al.* Scoping reviews: time for clarity in definition, methods, and reporting. *J. Clin. Epidemiol.* **67**, 1291–1294 (2014).
110. Arksey, H. & O'Malley, L. Scoping studies: towards a methodological framework. *Int. J. Soc. Res. Methodol.* **8**, 19–32 (2005).
111. Daudt, H. M., van Mossel, C. & Scott, S. J. Enhancing the scoping study

methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. *BMC Med. Res. Methodol.* **13**, 48 (2013).
112. Pham, M. T. *et al.* A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Res. Synth. Methods* **5**, 371–385 (2014).
113. Peters, M. D. J. *et al.* Guidance for conducting systematic scoping reviews. *Int. J. Evid. Based. Healthc.* **13**, 141–146 (2015).
114. Stewart, L. A. & Tierney, J. F. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval. Health Prof.* **25**, 76–97 (2002).
115. Bhandari, M., Morrow, F., Kulkarni, A. V & Tornetta, P. Meta-analyses in orthopaedic surgery: a systematic review of their methodologies. *J. Bone Joint Surg. Am.* **83–A**, 15–24 (2001).
116. Antman, E. M., Lau, J., Kupelnick, B., Mosteller, F. & Chalmers, T. C. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* **268**, 240–8 (1992).
117. Dickersin, K. & Min, Y. I. NIH clinical trials and publication bias. *Online J. Curr. Clin. Trials* **Doc No 50**, [4967 words; 53 paragraphs] (1993).
118. Ioannidis, J. P. A. Meta-research: The art of getting it wrong. *Res. Synth. Methods* **1**, 169–84 (2010).
119. Dwan, K., Gamble, C., Williamson, P. R., Kirkham, J. J. & Reporting Bias Group. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One* **8**, e66844 (2013).
120. Dickersin, K., Min, Y. I. & Meinert, C. L. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* **267**, 374–8 (1992).
121. Qunaj, L. *et al.* Delays in the publication of important clinical trial findings in oncology. *JAMA Oncol.* **4**, e180264 (2018).
122. Chalmers, T. C., Frank, C. S. & Reitman, D. Minimizing the three stages of publication bias. *JAMA* **263**, 1392–5 (1990).
123. Lexchin, J., Bero, L. A., Djulbegovic, B. & Clark, O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* **326**, 1167–70 (2003).
124. Heres, S. *et al.* Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am. J. Psychiatry* **163**, 185–94 (2006).
125. Melander, H., Ahlqvist-Rastad, J., Meijer, G. & Beermann, B. Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* **326**, 1171–3 (2003).
126. Battisti, W. P. *et al.* Good Publication Practice for Communicating Company-Sponsored Medical Research: GPP3. *Ann. Intern. Med.* **163**,

461–4 (2015).
127. ICMJE. Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication. *J. Pharmacol. Pharmacother.* **1**, 42–58 (2010).
128. Goldacre, B. *et al.* Compliance with requirement to report results on the EU Clinical Trials Register: cohort study and web resource. *BMJ* **362**, k3218 (2018).
129. Pautasso, M. Ten simple rules for writing a literature review. *PLoS Comput. Biol.* **9**, e1003149 (2013).
130. Ebrahim, S., Bance, S., Athale, A., Malachowski, C. & Ioannidis, J. P. A. Meta-analyses with industry involvement are massively published and report no caveats for antidepressants. *J. Clin. Epidemiol.* **70**, 155–63 (2016).
131. Booth, A. *et al.* The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Syst. Rev.* **1**, 2 (2012).
132. Cullis, P. S., Gudlaugsdottir, K. & Andrews, J. A systematic review of the quality of conduct and reporting of systematic reviews and meta-analyses in paediatric surgery. *PLoS One* **12**, e0175213 (2017).
133. Whiting, P., Westwood, M., Burke, M., Sterne, J. & Glanville, J. Systematic reviews of test accuracy should search a range of databases to identify primary studies. *J. Clin. Epidemiol.* **61**, 357–364 (2008).
134. Atkinson, L. Z. & Cipriani, A. How to carry out a literature search for a systematic review: a practical guide. *BJPsych Adv.* **24**, 74–82 (2018).
135. US National Library of Medicine. MEDLINE, PubMed and PMC (PubMed Central) - How are they different? *Fact Sheet* 1 (2016). doi:10.1002/ijc.28260
136. Embase. Embase Coverage and Content. *Elsevier* (2017). Available at: https://www.elsevier.com/solutions/embase-biomedical-research/embase-coverage-and-content. (Accessed: 19th August 2018)
137. Elsevier B.V. Scopus: Content Coverage Guide. 28 (2017).
138. Suarez-Almazor, M. E., Belseck, E., Homik, J., Dorgan, M. & Ramos-Remus, C. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. *Control. Clin. Trials* **21**, 476–87 (2000).
139. Rudmik, L. R., Walen, S. G., Dixon, E. & Dort, J. Evaluation of meta-analyses in the otolaryngological literature. *Otolaryngol. Neck Surg.* **139**, 187–194 (2008).
140. Preston, L., Carroll, C., Gardois, P., Paisley, S. & Kaltenthaler, E. Improving search efficiency for systematic reviews of diagnostic test accuracy: an exploratory study to assess the viability of limiting to MEDLINE, EMBASE and reference checking. *Syst. Rev.* **4**, 82 (2015).
141. Jadad, A. R. & McQuay, H. J. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. *J. Clin. Epidemiol.* **49**, 235–43 (1996).
142. Lacasse, J. R. & Leo, J. Ghostwriting at elite academic medical centers in

the United States. *PLoS Med.* **7**, e1000230 (2010).
143. Leopold, S. S., Warme, W. J., Fritz Braunlich, E. & Shott, S. Association between funding source and study outcome in orthopaedic research. *Clin. Orthop. Relat. Res.* 293–301 (2003). doi:10.1097/01.blo.0000093888.12372.d9
144. Perlis, R. H. *et al.* Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. *Am. J. Psychiatry* **162**, 1957–60 (2005).
145. Yaphe, J., Edman, R., Knishkowy, B. & Herman, J. The association between funding by commercial interests and study outcome in randomized controlled drug trials. *Fam. Pract.* **18**, 565–8 (2001).
146. Lundh, A., Lexchin, J., Mintzes, B., Schroll, J. B. & Bero, L. Industry sponsorship and research outcome: systematic review with meta-analysis. *Intensive Care Med.* **44**, 1603–1612 (2018).
147. Egger, M., Bartlett, C. & Jüni, P. Are randomised controlled trials in the BMJ different? *BMJ* **323**, 1253–4 (2001).
148. Tallon, D., Chard, J. & Dieppe, P. Relation between agendas of the research community and the research consumer. *Lancet (London, England)* **355**, 2037–40 (2000).
149. Gøtzsche, P. C. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control. Clin. Trials* **10**, 31–56 (1989).
150. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science (80-.).* **349**, aac4716-aac4716 (2015).
151. Leichsenring, F. *et al.* Biases in research: risk factors for non-replicability in psychotherapy and pharmacotherapy research. *Psychol. Med.* **47**, 1000–1011 (2017).
152. Hengartner, M. P. Raising Awareness for the Replication Crisis in Clinical Psychology by Focusing on Inconsistencies in Psychotherapy Research: How Much Can We Rely on Published Findings from Efficacy Trials? *Front. Psychol.* **9**, (2018).
153. Stanley, T. D., Carter, E. C. & Doucouliagos, H. What meta-analyses reveal about the replicability of psychological research. *Psychol. Bull.* **144**, 1325–1346 (2018).
154. Guyatt, G. H. *et al.* What is "quality of evidence" and why is it important to clinicians? *BMJ* **336**, 995–998 (2008).
155. Guyatt, G. H. *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336**, 924–926 (2008).
156. Delaney, A. *et al.* A systematic evaluation of the quality of meta-analyses in the critical care literature. *Crit. Care* **9**, R575-82 (2005).
157. Moher, D. PRISMA 2009 Checklist. *PLoS medicine* **6**, e1000097 (2009).
158. Tricco, A. C. *et al.* PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann. Intern. Med.* **169**, 467 (2018).
159. Barnes, D. E. & Bero, L. A. Why review articles on the health effects of passive smoking reach different conclusions. *JAMA* **279**, 1566–70

(1998).
160. Assendelft, W. J., Koes, B. W., Knipschild, P. G. & Bouter, L. M. The relationship between methodological quality and conclusions in reviews of spinal manipulation. *JAMA* **274**, 1942–8 (1995).
161. Sung, J. J. Y., Tsoi, K. K. F., Lai, L. H., Wu, J. C. Y. & Lau, J. Y. W. Endoscopic clipping versus injection and thermo-coagulation in the treatment of non-variceal upper gastrointestinal bleeding: a meta-analysis. *Gut* **56**, 1364–73 (2007).
162. Jairath, V. & Barkun, A. N. Nonvariceal upper GI bleeding: it's not just about peptic ulcers. *Gastrointest. Endosc.* **75**, 273–275 (2012).
163. Di Castelnuovo, A. Alcohol Dosing and Total Mortality in Men and Women. *Arch. Intern. Med.* **166**, 2437 (2006).
164. Bagnardi, V. Flexible Meta-Regression Functions for Modeling Aggregate Dose-Response Data, with an Application to Alcohol and Mortality. *Am. J. Epidemiol.* **159**, 1077–1086 (2004).
165. Holman, C. D., English, D. R., Milne, E. & Winter, M. G. Meta-analysis of alcohol and all-cause mortality: a validation of NHMRC recommendations. *Med. J. Aust.* **164**, 141–5 (1996).
166. Plunk, A. D., Syed-Mohammed, H., Cavazos-Rehg, P., Bierut, L. J. & Grucza, R. A. Alcohol consumption, heavy drinking, and mortality: rethinking the j-shaped curve. *Alcohol. Clin. Exp. Res.* **38**, 471–8 (2014).
167. Wang, C. *et al.* Effect of Drinking on All-Cause Mortality in Women Compared with Men: A Meta-Analysis. *J. Women's Heal.* **23**, 373–381 (2014).
168. Ronksley, P. E., Brien, S. E., Turner, B. J., Mukamal, K. J. & Ghali, W. A. Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. *BMJ* **342**, d671–d671 (2011).
169. Costanzo, S., Di Castelnuovo, A., Donati, M. B., Iacoviello, L. & de Gaetano, G. Alcohol Consumption and Mortality in Patients With Cardiovascular Disease. *J. Am. Coll. Cardiol.* **55**, 1339–1347 (2010).
170. Xi, B. *et al.* Relationship of Alcohol Consumption to All-Cause, Cardiovascular, and Cancer-Related Mortality in U.S. Adults. *J. Am. Coll. Cardiol.* **70**, 913–922 (2017).
171. Fekjaer, H. O. Alcohol-a universal preventive agent? A critical analysis. *Addiction* **108**, 2051–2057 (2013).
172. REHM, J. *et al.* Alcohol as a risk factor for liver cirrhosis: A systematic review and meta-analysis. *Drug Alcohol Rev.* **29**, 437–445 (2010).
173. Goulden, R. Moderate Alcohol Consumption Is Not Associated with Reduced All-cause Mortality. *Am. J. Med.* **129**, 180–186.e4 (2016).
174. Fillmore, K. M., Stockwell, T., Chikritzhs, T., Bostrom, A. & Kerr, W. Moderate Alcohol Use and Reduced Mortality Risk: Systematic Error in Prospective Studies and New Hypotheses. *Ann. Epidemiol.* **17**, S16–S23 (2007).
175. Stockwell, T. *et al.* Alcohol-caused mortality in australia and Canada:

scenario analyses using different assumptions about cardiac benefit. *J. Stud. Alcohol Drugs* **68**, 345–52 (2007).
176. Stockwell, T. *et al.* Do 'Moderate' Drinkers Have Reduced Mortality Risk? A Systematic Review and Meta-Analysis of Alcohol Consumption and All-Cause Mortality. *J. Stud. Alcohol Drugs* **77**, 185–98 (2016).
177. Nih.gov. NIH to end funding for Moderate Alcohol and Cardiovascular Health trial. *NEWS RELEASES* (2018). Available at: https://www.nih.gov/news-events/news-releases/nih-end-funding-moderate-alcohol-cardiovascular-health-trial. (Accessed: 7th January 2019)
178. Ferrara, R., Mezquita, L. & Besse, B. Progress in the Management of Advanced Thoracic Malignancies in 2017. *J. Thorac. Oncol.* **13**, 301–322 (2018).
179. Dutch Anti-Doping Authority. *Annual Report*. **8**, (2013).
180. International Society for Pharmacoeonomics and Outcomes Research. ISPOR and ISPE Collaborate to Advance Good Practices for Use of Real-World Evidence. (2017). Available at: https://press.ispor.org/index.php/ispor-and-ispe-collaborate-to-advance-good-practices-for-use-of-real-world-evidence/. (Accessed: 4th October 2019)
181. Moher, D., Tetzlaff, J., Tricco, A. C., Sampson, M. & Altman, D. G. Epidemiology and Reporting Characteristics of Systematic Reviews. *PLoS Med.* **4**, e78 (2007).
182. Ioannidis, J. P. A., Chang, C. Q., Lam, T. K., Schully, S. D. & Khoury, M. J. The Geometric Increase in Meta-Analyses from China in the Genomic Era. *PLoS One* **8**, e65602 (2013).
183. Seife, C. For sale: "Your Name Here" in a prestigious science journal. *Scientific American* (2014). Available at: https://www.scientificamerican.com/article/for-sale-your-name-here-in-a-prestigious-science-journal/.
184. Dixon, E., Hameed, M., Sutherland, F., Cook, D. J. & Doig, C. Evaluating meta-analyses in the general surgical literature: a critical appraisal. *Ann. Surg.* **241**, 450–9 (2005).
185. Oxman, A. D. & Guyatt, G. H. Validation of an index of the quality of review articles. *J. Clin. Epidemiol.* **44**, 1271–8 (1991).
186. Cooney, G. M. *et al.* Exercise for depression. *Cochrane Database Syst. Rev.* (2013). doi:10.1002/14651858.CD004366.pub6
187. Tsafnat, G. *et al.* Systematic review automation technologies. *Syst. Rev.* **3**, 1–15 (2014).
188. Memorial Sloan Kettering. Memorial Sloan Kettering Trains IBM Watson to Help Doctors Make Better Cancer Treatment Choices. (2014).
189. Ross, C. & Swetlitz, I. IBM's Watson recommended 'unsafe and incorrect' cancer treatments - STAT. *STAT* (2018).
190. Horder, T. J. The organizer concept and modern embryology: Anglo-American perspectives. *Int. J. Dev. Biol.* **45**, 97–132 (2001).

11 Abstract

Evidence-based medicine (EBM) is the integration of the best available evidence with the specific details of a patient, to select the best treatment for that individual. To do this, healthcare providers (HCPs) need to have access to the best evidence at hand. Systematic reviews (SRs) help by making that information available.

SRs use a specific methodology, with a systematic process to find, appraise and collate all relevant evidence to answer a research question. This helps HCPs to select the best option for their patient. Though SRs are common in EBM, they are not as common in other health-related questions, though they could be useful. To do this, several aspects of this SR methodology need to be tailored to make sure it is fit-for-purpose.

For example, the search methodology needs to be changed. Where an SR in EBM would use a thesaurus, which is a structured list of terms, to find all the relevant publications with a specific disease, this is harder for questions that do not focus on a specific disease or treatment. Different databases should also be considered to find the relevant information. Finally, the standard methods of compiling data with forest plots and summary statistics are not always possible. These choices all must be considered when designing an SR on a health-related topic outside EBM.

This thesis contextualises the choices in the design of such an SR and evaluates the appropriateness of each method for a specific research question, so that the overall research question can be answered: What choices and considerations need to be made in the design of an SR for a health-related research question, and how do they impact the outcome and conclusion? We discuss five different research questions, and tailor the methodology of an SR to find the most reliable answer to each question.

First, looking at what risk factors impact criminality, we chose to formulate a search query using the thesauri of search engines. Within EBM, a researcher can assume that the search engine uses standard language and indexing, which is less reliable outside EBM. Going through each relevant term allowed us to select exactly which terms we wanted to include, so that we could retrieve all relevant records.

We debate whether this approach saved us time, as we still found many irrelevant records. For future SRs, it might not be worthwhile to decide for each term if it should be included in the search or not. This chapter has shown that it is not essential to go through a thesaurus in detail to find all relevant information. In terms of content, this SR yielded interdisciplinary insights into criminology: criminality is the end product of a complex chain of genes, risks and exposures. A multidisciplinary, individualized approach remains necessary for forensic risk-assessments.

Second, researching the reasons for the recent outcry about increases in drug pricing, we searched sources from both scientific and newspaper databases to answer a question that spans both domains. We used four newspapers: two British and two American, two with a general and two with a business focus.

Tailoring the search methodology to both scientific and newspaper search engines allowed us to analyse both the scientific debate and the public outcry, with different nuances. This approach added value by illuminating various sides of the arguments. Whereas the public debate was mainly focussed on affordability of drugs for individuals, and gave examples of poor outcomes at the individual level, the scientific literature focussed on public health outcomes, policy proposals and their effects, with abstract calculations in quality-adjusted life years, which frequently leads to opposing opinions. We offer an overview of policy interventions to improve access and prices and their outcomes, so that policy makers can consider what the most appropriate solution is in their setting. Our expertise on this area gave us enough credibility to engage in the academic debate by writing letters critiquing the findings of other authors.

Third, we estimated the potential of gene doping use of specific proteins, by using a standard search string which was adjusted for each gene. Gene doping is the use of a method to change genetic information for better achievements in sport, comparable to gene therapy for patients. We considered how to combine the findings and quantify a qualitative field, so that we could choose which compounds were most promising for abuse.

We made an overview of the mechanisms of and experience with each gene, the properties of the most common vectors and how gene

doping would be detected. After that we developed a scoring mechanism, which allowed us to judge each gene for its attributes that make it likely to be abused and develop a prediction on which one would most likely be abused. EPO and VEGF came out as most likely candidates. The individually modified search string yielded good results. This framework and search strategy can be used for other health-related research questions too.

Our fourth question considered the reliability of findings in a quantitative research area. Addressing the effect of three specific fibres on cholesterol in animal research, we compared the findings for each fibre by plotting it in a graph and calculating an overall effect on different types of cholesterol. After that, we produced a funnel plot, so that we could look at the reliability of the findings.

Here, we found publication bias, meaning selective reporting depending on whether the results were deemed worth publishing, and considered how to best use the findings. This shows that publication bias is also present in research areas outside EBM and can be demonstrated with similar tools to those used in EBM. There are some downsides to using data from different animals and different study methods, but the overall findings are convincing. We calculated that chitosan has the best hypocholesterolaemic profile of the three selected fibres. We also mapped the mechanisms behind the effect of the fibres. Finally, we considered if a designer fibre, with the best mechanistical properties of each of the fibres we researched, would be possible.

Finally, researching depression in adolescents, we considered whether a narrative review was not more appropriate then an SR. As many reviews had already been written, an extra review would not have helped the science forward. The field needed a breakthrough, though, so we considered a different approach. We initially stepped away from the literature and started from the patient perspective to try to force a breakthrough in thinking about the treatment of depression.

This was not an SR, because that would not have been fit for the purpose of formulating a new hypothesis. This approach led us to focus more on the patients' stress and coping behaviour, and less on symptoms and drugs. If independently validated, this could cause a paradigm shift of the treatment of depression in clinical practice.

Each case study uses a different method, demonstrating that there are many opportunities to tailor an SR. SR is a versatile tool, which, if used properly, can address many different research questions. The search strategy can be changed to include the most relevant search engines and search terms. The method of compiling data needs consideration, but if done well can add a clear visual that summarizes all included studies. The high number of citations and downloads of our work, and of SRs in general, show that other researchers view it as a reliable overview and deem it valuable for their research field, which underline the value of this tool to science. Optimizing the methodology of an SR maximises the chances of finding the most reliable estimate to answer a health-related research question.

The landscape of SRs is likely to face some challenges going forward. Scandals with authorship for sale and industry-funded SRs show the lack of quality in peer-review. Methodologically poor reviews are often published anyway, either intentionally or unintentionally shaping the scientific debate. With the development of more tools to facilitate or even automate steps in the writing of SRs, such as referencing software and analysis tools, it will be easier to write methodologically bad reviews, so it will be even more important to audit the methodology of SRs.

Specifically for SRs on health-related research questions outside EBM, where the methodology is less standardised, a good understanding of the methodology of the SR is essential. The databases, search queries, and method of compilations need to be tailored to fit the question and research goal.

In conclusion, our works shows that performing SRs in health-related research outside evidence-based medicine requires careful consideration of all steps in the methodology. If done well, SRs accelerate the progress and implementation of scientific knowledge.

12 Samenvatting

Evidence-based medicine (EBM) is de integratie van het beste beschikbare bewijs met de specifieke details van een patiënt, om zo de beste mogelijke behandeling voor een specifiek individu te kiezen. Om dit te doen, moeten zorgverleners (ZV) toegang hebben tot het beste bewijs. Systematic reviews (SRs) helpen door die informatie beschikbaar te maken.

SRs gebruiken een specifieke methodologie, met een systematisch proces om al het relevante bewijs om een onderzoeksvraag te beantwoorden te vinden, te beoordelen en samen te voegen. Dit helpt ZVs om de beste optie voor hun patiënt te vinden. Hoewel SRs vaak gebruikt worden in EBM, komen ze minder vaak voor in ander gezondheidsonderzoek, terwijl ze daar ook zinvol kunnen zijn. Om SRs hier toe te passen moeten verschillende aspecten van de methodologie worden aangepast om te zorgen dat die adequaat is voor het doel

De zoekmethodologie moet bijvoorbeeld worden veranderd. Waar een SR in EBM een thesaurus (een gestructureerde lijst van termen) kan gebruiken om alle relevante publicaties voor een specifieke ziekte te vinden, is dit lastiger voor vragen die niet op een ziekte of behandeling focussen. Verschillende databases moeten ook worden overwogen om alle informatie te vinden. En tenslotte zijn de standaardmethodes om data samen te voegen, met forest plots en samenvattende statistieken, niet altijd mogelijk. Deze keuzes moeten allemaal worden overwogen in het ontwerp van een SR over een gezondheidsgerelateerd onderwerp buiten EBM.

Deze thesis contextualiseert de keuzes in het ontwerp van een SR en evalueert de toepasselijkheid van elke methode voor de gezondheidsgerelateerde onderzoeksvraag, zodat we de hoofdonderzoeksvraag kunnen beantwoorden: welke keuzes en overwegingen moeten er worden gemaakt bij het ontwerpen van een SR voor een gezondheidsgerelateerde onderzoeksvraag, en hoe beïnvloeden die de uitkomsten en conclusie? We bestuderen vijf onderzoeksvragen, en passen de methodologie daarop aan om de meest relevante literatuur te vinden en beoordelen.

Ten eerste kijken we naar risicofactoren in de criminologie, waarbij we de zoektermen formuleren aan de hand van de thesauri van de zoekmachines. Binnen EBM kan een onderzoeker aannemen dat de

zoekmachines gestandaardiseerde taal en indexering gebruiken, maar dat is minder betrouwbaar buiten EBM. Door alle relevante termen te selecteren konden we precies selecteren wat we wilden includeren, zodat we alle relevante literatuur konden vinden.

Of dit ons tijd heeft bespaard is onzeker, aangezien we nog steeds veel irrelevante literatuur vonden. Voor toekomstige SRs is het niet altijd de moeite waard om voor elke term door te nemen of het geïncludeerd moet worden in de zoekvraag. Dit hoofdstuk heeft duidelijk gemaakt dat het niet noodzakelijk is om een thesaurus gedetailleerd door te nemen om relevante informatie te vinden. Inhoudelijk heeft deze SR interdisciplinaire inzichten in de criminologie opgeleverd: criminaliteit is het eindproduct van genen, risicofactoren en blootstellingen. Een multidisciplinaire benadering op het individuele niveau blijft noodzakelijk bij forensische risicoanalyses.

Ten tweede onderzochten we het recente debat over de stijgende medicijnprijzen, waarbij we bronnen van zowel wetenschappelijke literatuur als kranten includeerden, zodat we een vraag kunnen beantwoorden die zowel het medische als publieke domein betreft. We gebruikten vier kranten: twee Britse en twee Amerikaanse, twee met een algemene en twee met een financiële focus.

Door de zoekmethodologie op zowel wetenschappelijke als zoekmachines van kranten aan te passen konden we het wetenschappelijke debat en publieke debat analyseren, met verschillende nuances. Deze benadering verhelderde verschillende aspecten die een rol spelen in het debat over medicijnkosten. Waar het publieke debat vooral focust op betaalbaarheid van medicijnen voor individuele patiënten, en vaak enkele voorbeelden op menselijke schaal gaf, gaat het in de wetenschappelijke literatuur veel meer over volksgezondheid en beleid, met abstracte berekeningen per quality-adjusted life years, wat vaak tot conflicterende meningen leidt. We bieden een overzicht van beleidsmogelijkheden om toegang tot medicijnen en de prijzen te verbeteren, zodat beleidsmakers kunnen overwegen wat een goede oplossing in hun markt is. Onze expertise over dit onderwerp gaf ons de geloofwaardigheid om in het academische debat mee te doen, door brieven te schrijven waarin we de bevindingen van andere auteurs beoordeelden.

Ten derde onderzochten we de mogelijkheid om specifieke eiwitten te gebruiken voor gendoping, door een standaard zoekvraag te gebruiken en die aan te passen voor elk gen. Gendoping is het gebruik van een methode om genetische informatie aan te passen voor betere sportprestaties, vergelijkbaar met gentherapie voor patiënten. We overwogen hoe we de bevindingen konden combineren en hoe we dit kwalitatieve veld konden kwantificeren om te beslissen welk stofje het meest waarschijnlijk als eerst misbruikt zou worden.

We presenteerden een overzicht van de mechanismes van en ervaring met elk gen, de eigenschappen die de meest typische vectoren hebben en hoe gendoping opgespoord zou kunnen worden. Daarna ontwierpen we een scoringsmechanisme, waardoor we elk gen op zijn eigenschappen konden beoordelen, zodat we een voorspelling konden doen over welk gen het meest kans maakt om als gendoping misbruikt te worden. EPO en VEGF kwamen als meest waarschijnlijke kandidaten naar voren. De zoekmethode met aanpassingen per gen leverde goede resultaten op. Deze opzet en zoekstrategie kan ook voor andere gezondheidsgerelateerde onderzoeksvragen gebruikt worden.

Onze vierde vraag ging over de betrouwbaarheid van resultaten in een kwantitatief onderzoeksveld. We keken naar de effecten van drie vezels op cholesterol in dieronderzoek, en vergeleken de bevindingen door het in een grafiek te tekenen en een totaal effect per type cholesterol te berekenen. Daarna maakten we een funnel plot, zodat we de betrouwbaarheid van de bevindingen konden bekijken.

Hier vonden we publicatiebias, waarbij resultaten selectief worden gepubliceerd op basis van of de resultaten het publiceren waard zijn, en we overwogen hoe we de resultaten het best konden gebruiken. Dit laat zien dat publicatiebias ook in onderzoeksvelden buiten EBM bestaat en aan te tonen is met vergelijkbare methodes als gangbaar is binnen EBM. Er zijn nadelen aan het gebruiken van data van verschillende diersoorten en verschillende onderzoeksmethoden, maar het uiteindelijke resultaat is overtuigend. We berekenden dat chitosan het beste cholesterolverlagende profiel heeft van de drie geselecteerde vezels. Ook hebben we de mechanismen van het effect van deze vezels in kaart gebracht. Tenslotte bespraken we het mogelijke ontwerp van een nieuwe vezel, met de beste eigenschappen van deze vezels die we onderzocht hadden.

Ten slotte keken we naar depressie in adolescenten, en overwogen we dat een narrative review toepasselijker was dan een SR. Er waren al zo veel SRs geschreven dat een extra review geen stap vooruit zou kunnen betekenen voor de wetenschap. Het veld had een doorbraak nodig, dus bedachten we een andere aanpak. Initieel lieten we de literatuur voor wat het was en begonnen we met het perspectief van de patiënten, zodat we hopelijk een doorbraak zouden kunnen forceren.

Dit was geen SR, omdat dat niet gepast was voor het doel om een nieuwe hypothese te formuleren. Deze aanpak zorgde er voor dat we op stress en mechanismes om daar mee om te gaan bij patiënten focusten, en minder op symptomen en medicijnen. Als dit onafhankelijk gevalideerd wordt, kunnen deze inzichten een verschuiving van het behandelparadigma voor depressie in de klinische praktijk teweegbrengen.

Elke casus gebruikte verschillende methodes, wat aantoont dat er veel mogelijkheden zijn om SRs aan te passen. SR is een veelzijdig gereedschap, dat, als het goed gebruikt wordt, veel verschillende onderzoeksvragen kan onderzoeken. De zoekstrategie kan worden aangepast om de meest relevante zoekmachines en zoektermen te includeren. De methode om data samen te voegen moet goed overwogen worden, maar kan een duidelijk visueel overzicht geven van alle studies als het goed wordt gedaan. De vele citaties en downloads van ons werk, en van SRs in het algemeen, laat zien dat andere onderzoekers het een betrouwbaar overzicht vinden en het waardevol achten voor hun veld van onderzoek, wat de waarde van SRs voor de wetenschap laat zien. De methodologie van een SR optimaliseren maximaliseert de kans dat het meest betrouwbare antwoord op een gezondheidsgerelateerde onderzoeksvraag wordt gevonden.

Het landschap van SRs gaat waarschijnlijk een aantal bedreigingen zien in de komende jaren. Schandalen met te koop aangeboden auteurschap en door de industrie betaalde SRs laten het gebrek aan kwaliteit in peer review zien. Methodologisch zwakke SRs worden vaak toch gepubliceerd, waarbij het wetenschappelijke debat bedoeld of onbedoeld wordt beïnvloed. Met de ontwikkeling van nieuwe gereedschappen om het schrijven van SRs te faciliteren of zelfs automatiseren, zoals referentiesoftware en analysegereedschappen, wordt het makkelijker om methodologisch slechte reviews te schrijven,

waardoor het nog belangrijker wordt om goed naar de methode van een SR te kijken.

Specifiek voor SRs in gezondheidsgerelateerde onderzoeksvragen buiten EBM, waar de methodologie minder gestandaardiseerd is, is het belangrijk de methode van een SR goed te begrijpen. De databases, de zoekstrategie en de methode van het samenvoegen moeten aangepast worden om bij de zoekvraag en het doel te passen.

Ter conclusie, ons werk laat zien dat het uitvoeren van SRs in gezondheidsgerelateerde onderzoeksgebieden buiten EBM uitgebreide aanpassing aan alle stappen in de methodologie vereist. Als het goed gedaan wordt, versnellen SRs de voortgang en implementatie van wetenschappelijke kennis.

13 Plain language abstract

In evidence-based medicine, systematic reviews (SRs) are often performed to determine the best treatment. This methodology is a systematic process to find and collate all the relevant evidence to answer a research question. Researchers in other health-related questions don't often use this methodology, though this could be useful. In this thesis, we explore how the SR method needs to be tailored to make it most useful for health-related research questions.

We look at five research questions and adjusted the methodology to best fit the question. First, we use thesauri (structured lists of search terms) to select exactly which search terms we want to use, to find relevant literature about risk factors for criminality. Second, we looked at what scientific journals and newspapers wrote about drug prices, to get a broad perspective. Because of our work and our experience in drug pricing, we could respond to other published papers. Third, we looked at how to score proteins in gene doping, so that we could predict which one would be most likely to be abused. Fourth, we looked at cholesterol-lowering fibres in animal research, and found publication bias. Finally, we wrote a narrative review on depression in adolescents, because that was more appropriate than an SR to reach a breakthrough in depression treatment.

In conclusion, to write good SRs on health-related research, extensive tailoring of the methodology is required. This requires expertise in SR methodology, but adds high value to scientific progress.

14 Samenvatting in gewoon Nederlands

In evidence-based medicine (EBM) worden systematic reviews (SRs) vaak gebruikt om vast te stellen wat de beste behandeling is. Deze methodologie is een systematisch proces om al het relevante bewijs te vinden en samen te voegen om een onderzoeksvraag te beantwoorden. Onderzoekers in andere gezondheidsgerelateerde onderzoeksvragen gebruiken deze methodologie minder vaak, terwijl dit zou kunnen helpen. In deze thesis onderzoeken we hoe de methodologie van SRs moet worden aangepast om het zinvoller te maken voor gezondheidsgerelateerde onderzoeksvragen.

We keken naar vijf onderzoeksvragen en pasten de methodologie aan de vraag aan. Ten eerste gebruikten we thesaurussen (gestructureerde lijsten van zoektermen) om precies te selecteren waar we op wilden zoeken, zodat we relevante literatuur over risicofactoren in criminaliteit konden vinden. Ten tweede onderzochten we wat wetenschappelijke tijdschriften en kranten over medicijnprijzen schreven, zodat we een breed perspectief kregen. Door ons werk en onze expertise konden we reageren op andere publicaties. Ten derde keken we naar hoe eiwitten die kunnen worden misbruikt voor gendoping gescoord kunnen worden, zodat we kunnen voorspellen welke het meest waarschijnlijk misbruikt wordt. Ten vierde keken we naar cholesterolverlagende vezels in dieronderzoek, waar we publicatiebias vonden. Ten slotte schreven we een narrative review over depressie in adolescenten, omdat dat beter paste dan een SR om een doorbraak in de behandeling te bereiken.

Ter conclusie vereist het schrijven van SR expertise in de methodologie en in de inhoud, maar het voegt veel waarde toe aan wetenschappelijke vooruitgang. Om goede SRs te schrijven in gezondheidsgerelateerd onderzoek is uitgebreide aanpassing van de methodologie nodig.

15 Acknowledgements

The first thanks goes to Toine Pieters, who guided me through this process. By providing me with new, interesting research topics he kept me motivated to dig deeper into the methodology. He gave me the autonomy to consider the best options, but was involved in every step of the way to ensure the final product was valuable for the scientific community.

I would also like to thank my sister, Marie-Christine, who kept on pushing me to achieve more and finalize this thesis, and my brother, Bonheur, who also stimulated me to push through. My parents, Ton and Gemma, who supported me during my studies and allowed me to develop scientifically, also deserve acknowledgements.

In the last couple of years, I met several people who broadened my horizons with new research projects, whom I would like to thank. First Carlo Bertucci and Edoardo Fabini, who guided me during my internship in at the university of Bologna and offered friendship and scientific debate. Rogier Lange, for showing me that good science and small-scale research can have a significant impact on patients' lives. I would also like to thank Jelle Sterk, who has showed interest in all my work and was willing to discuss it even when it was probably not interesting anymore.

In Malawi, I worked with the cancer program team of UNC Lilongwe, who taught me so much. Thank you, Satish Gopal, for giving me the opportunity, and thank you team, Edwards Kasonkanji, Bongani Kaimila, Christopher Stanley, Tamiwe Tomoka and Robin Kajasiche, for all your insights, help, friendship and an unforgettable experience. Through your examples I learned about the value of observational research in evidence generation, and the use of various research tools in a challenging setting.

And most importantly, I would like to thank Cassandra Nemzoff, with whom I discussed many findings, who proofread almost all my publications and provided strict yet fair feedback, and who provided so much encouragement in the last couple of years. I wouldn't have done this without you.

16 CV

16.1 Education

PhD at Utrecht University

Nov 2013 – Dec 2019

Thesis: Systematic Review Methodology in Biomedical Evidence Generation

Master Pharmaceutical Sciences (MSc/PharmD) at Utrecht University

Feb 2012 - Nov 2015

GPA: 4.0. Not currently registered as pharmacist.

Bachelor Pharmaceutical Sciences (BSc) at Utrecht University

Sept 2008 - Sept 2012

GPA: 4.0; Best thesis of my cohort. 2007-2008: Minor Mathematics (probability & statistics, algebra, calculus)

16.2 Professional experience

16.2.1 Clinical scientist

Apr 2019 - current

AstraZeneca, Global Medicine Development, Gmed Oncology, Tagrisso (Osimertinib)

Cambridge, UK

- Contributed to the development of clinical study protocols, documents and rationale
- Contributed to clinical interpretation of data and future research directions
- Provided safety and data integrity oversight during the study conduct
- Supported decision making into ongoing and future studies
- Conducted training to internal and external audiences

16.2.2 Global publication specialist

Feb 2018 - Mar 2019

AstraZeneca, Global Medical Affairs, Oncology Business Unit

Cambridge, UK

- Led the publication of weekly oncology literature reports, improved search and reporting methodology, evaluated strategic impact and relevance to AZ and communicated to 5000+ worldwide AZ staff
- Reported on publication productivity of internal and external research publications to senior leadership
- Responsible for execution of $100k vendor contracts
- Liaised with siloed medical, clinical and publication staff, within AZ and external, to align and improve publication processes and ensure strategic delivery
- Trained internal staff and agencies on use of publication systems and audited compliance

16.2.3 Clinical Research Coordinator / Pharmaceutical Consultant

Apr 2016 – May 2017

University of North Carolina at Chapel Hill, Project-Malawi, Cancer Program

Lilongwe, Malawi

- Coordinated design, implementation, amendments and IRB approvals of clinical study protocols and ICFs compliant with GCP, the Declaration of Helsinki and HIPAA (ph1/dose escalation, ph2 on rituximab for local registration, observational cohorts, mixed methods)
- Provided literature and pharmacological rationale in study background, operationalisation and endpoints
- Led clinical study database design and functionality, performed QA and QC and rational use
- Merged datasets to answer specific research questions, completed statistical data-analyses and interpretation of multiple studies using MS Excel, Nvivo and Stata

- Supervised drafting and peer reviewed 25+ abstracts, posters and manuscripts to enhance accuracy and impact of research on cancers (mainly on Lymphomas, Breast and Ovarian cancer, Kaposi Sarcoma)
- Increased the number of annual primary publications relative to previous years by improving productivity and quality (2015: 4, 2016: 8, 2017: 17). See publications
- Reduced medication errors and improved quality of care by developing quality checks, customizing guidelines and redesigning prescription logistics
- Trained two nurses in safe chemotherapy preparation
- Lectured at University of Malawi, College of Medicine

16.2.4 Medical Writer

Apr 2012 – Mar 2016

Utrecht Institute for Pharmaceutical Sciences, Utrecht University

Utrecht, the Netherlands

- Wrote five systematic reviews and two letters to editors with Prof. Dr. Toine Pieters and experts.

16.2.5 Public Health Consultant

Aug 2015 – Sep 2015

GlaxoSmithKline, Care Solutions Department

Zeist, the Netherlands

- Led development of reports on impact of pharmaceutical policy changes, Asthma/COPD guideline changes and strategies to anticipate them, and implications of insurance risk pooling. Liaised with key stakeholders.

16.2.6 Postgraduate Scientist

Sept 2013 – Mar 2014

Università di Bologna, Pharmacy Faculty

Bologna, Italy

- Performed and analysed pharmacokinetic translational research on binding phenomena of bicalutamide analogues using surface

plasmon resonance and HPLC with immobilized human and rat serum albumin.

16.2.7 Pharmacy Intern

Feb 2013 – Jul 2015

- Gained experience in two public and one hospital pharmacy for 5 to 11 weeks each. Various projects, mainly in oncology.

16.3 Extracurricular activities

Mental Health England representative, Central Cambridge, AstraZeneca

Nov 2018 - current

Scouting Utrecht Oost, Boy Scout staff group chair

Feb 2007 - Sept 2015

- Managed and coordinated several official committees for scouting association. Organized and led activities, including weekend trips and summer camps, for 10 to 150 participants

Mega International Tournament (MIT) Chairman

Oct 2010 - Jun 2011

- Exercised project management skills to organize and execute the largest basketball tournament in the Benelux region with 500 players from across Europe

Other interests

- Languages: Dutch (native), English (fluent), Italian (advanced), German (basic), French (basic)
- Basketball, ultimate Frisbee and rugby player, avid guitar player

16.4 Bibliography

For an up to date overview, visit ORCID or Google scholar

16.4.1 Publications in peer-reviewed journals

- T van der Gronde, HG Leufkens, T Pieters. Response to proposal for a novel cancer drug pricing model. *Nature Reviews Clinical Oncology* 2018 Aug;15(8):528

- Breukels O, Van der Gronde T, Simons-Sanders K, Crul M. Antineoplastic Drug Contamination on the Outside of Prepared Infusion Bags. *Int J Pharm Comp*d. 2018 Jul-Aug;
- Westmoreland K, Reeve B, Amuquandoh A, van der Gronde T, Manthalu O, Correia H, Stanley C, Itimu S, Salima A, Chikasema M, Ward P, Mpasa A, Wachepa S, Mtete I, Butia M, Chasela M, Mtunda M, Wasswa P, Martin S, Kim N, Kazembe P, Gopal S. Translation, psychometric validation, and baseline results of the Patient-Reported Outcomes Measurement Information System (PROMIS) pediatric measures to assess health-related quality of life of patients with pediatric lymphoma in Malawi. *Pediatr Blood Cancer*. 2018 Jul 17:e27353
- T van der Gronde, T Pieters, Assessing Pharmaceutical Research and Development Costs, *JAMA internal medicine* 178 (4), 587-588
- Stanley, CC, Van der Gronde, T, Westmoreland, KD, Salima, A, Amuquandoh, A, Itimu, S, Manda, A, Mtete, I, Butia, M, Mpasa, A, Wachepa, S, Fox, P, Wasswa, P, Kazembe, P, El-Mallawany, NK, Gopal, S. Risk factors and reasons for loss to follow-up among children with lymphoma in Malawi. *Supportive Care in Cancer* 2017 1-7
- Tomoka, T, Powers, E, Van der Gronde, T, Amuquandoh, A, Dhungel, B, Kampani, C, Kamiza, S, Montgomery, N, Fedoriw, Y, Gopal, S. Extranodal Natural Killer/T-cell Lymphoma in Malawi: a report of three cases. *BMC Cancer* 2017 17:633
- Van der Gronde, T, Uyl-de Groot, CA, Pieters, T. Addressing the challenge of high-priced prescription drugs in the era of precision medicine: a systematic review of drug life cycles, therapeutic drug markets and regulatory frameworks. *PLoS ONE* 12(8): e0182613
- Kaimila B, Van der Gronde T, Stanley CC, Kasonkanji EE, Chikasema M, Tewete B, Fox P, Gopal S. Salvage chemotherapy for adults with relapsed or refractory lymphoma in Malawi. *Infectious Agents and Cancer* 2017; 12:45
- Kaimila B, van der Gronde T, Kasonkanji E, Fox P, Chikasema M, Tewete B, Gopal S. CD4 count and HIV RNA trends for HIV-associated lymphoproliferative disorders in Malawi. *AIDS Res Hum Retroviruses* 2017 Jun 26
- Westmoreland KD, Montgomery ND, Stanley CC, El-Mallawany NK, Wasswa P, van der Gronde T, Mtete I, Butia M, Itimu S, Chasela M,

Mtunda M, Kampani C, Liomba NG, Tomoka T, Dhungel BM, Sander MK, Krysiak R, Kazembe P, Dittmer DP, Fedoriw Y, Gopal S. Plasma Epstein-Barr virus DNA for pediatric Burkitt lymphoma diagnosis, prognosis, and response assessment in Malawi. *Int J Cancer* 2017 Mar 7

- Kasonkanji E, Kaimila B, Amuquandoh A, Chikasema M, Tewete B, Van der Gronde T, Gopal S. Successful treatment of classical Hodgkin lymphoma during pregnancy in Malawi. *J Oncol Pract* 2017 Feb 7
- Host KM, Horner MJ, Van der Gronde T, Moses A, Phiri S, Dittmer DP, Damania B, Gopal S. Kaposi's sarcoma in Malawi: a continued problem for HIV-positive and HIV-negative individuals. *AIDS* 2017 Jan 14;31(2):318-319
- Stanley CC, Westmoreland KC, Itimu S, Salima A, Van der Gronde T, Wasswa P, Mtete I, Butia M, El-Mallawany NK, Gopal S. Quantifying bias in survival estimates resulting from loss to follow-up among children with lymphoma in Malawi. *Pediatr Blood Cancer* 2016 Nov 28.
- Lange R, ter Heine R, Van der Gronde T, Selles S, de Klerk J, Bloemendal H, Hendrikse H. Applying quality by design principles to the small-scale preparation of the bone-targeting therapeutic radiopharmaceutical Rhenium-188-HEDP. *Eur J Pharm Sci* 2016 Jul 30;90:96-101
- Van der Gronde T, Hartog A, van Hees C, Pellikaan H, Pieters T. Systematic review on the evidence and mechanisms behind the hypocholesterolaemic effect of HPMC, pectin and chitosan in animal trials. *Food Chem* 2016;199:746-59
- Fortugno C, Van der Gronde T, Varchi G, Guerrini A, Bertucci C. Species-dependent binding of new synthesized bicalutamide analogues to albumin by optical biosensor analysis. *J Pharm Biomed Anal* 2015; 111:324-32
- Van der Gronde T, Kempes M, van El C, Rinne T, Pieters T. Neurobiological correlates in forensic assessment: a systematic review. *PLoS One* 2014 Oct 20;9(10)
- Van der Gronde T, de Hon O, Haisma HJ, Pieters T. Gene doping: an overview and current implications for athletes. *Br J Sports Med* 2013 Jul;47(11):670-8

16.4.2 Conference Abstracts

- 7, including SIOP, ASH, and AMP. See Google scholar

16.4.3 Invited talks

- Chemotherapy preparation in a Class III Biological Safety Cabinet. 21 October 2016, Baylor College of Medicine Children's Foundation Malawi, Lilongwe, Malawi
- Pharmacodynamics of antineoplastic drugs – principles of antiproliferative therapy in cancer. 2 November 2016; 7 December 2016, 2 March 2017, and 4 May 2017, University of Malawi, College of Medicine, Lilongwe, Malawi
- Cancer and HIV: basics and epidemiology. 2 March 2017, University of Malawi, College of Medicine, Lilongwe, Malawi

16.4.4 Clinical trial protocols

- A single arm phase I/II clinical trial of COMP with high dose methotrexate chemotherapy for Burkitt Lymphoma in Malawi. PI: Kate Westmoreland, MD. Principal co-PI: Satish Gopal, MD, MPH. Co-PIs: Toon van der Gronde, PharmD, Patrick Thompson, MD, Danny Gonzalez, PharmD, Nader El-Mallawany Kim, MD

17 FI Scientific Library
(formerly published as CD-β Scientific Library)

101. Klein, W. (2018). *New Drugs for the Dutch Republic. The Commodification of Fever Remedies in the Netherlands (c. 1650-1800).*
100. Flis, I. (2018). *Discipline Through Method - Recent history and philosophy of scientific psychology (1950-2018).*
99. Hoeneveld, F. (2018). *Een vinger in de Amerikaanse pap. Fundamenteel fysisch en defensie onderzoek in Nederland tijdens de vroege Koude Oorlog.*
98. Stubbé-Albers, H. (2018). *Designing learning opportunities for the hardest to reach: Game-based mathematics learning for out-of-school children in Sudan*.
97. Dijk, G. van (2018). *Het opleiden van taalbewuste docenten natuurkunde, scheikunde en techniek: Een ontwerpgericht onderzoek.*
96. Zhao, Xiaoyan (2018). *Classroom assessment in Chinese primary school mathematics education.*
95. Laan, S. van der (2017). *Een varken voor iedereen. De modernisering van de Nederlandse varkensfokkerij in de twintigste eeuw.*
94. Vis, C. (2017). *Strengthening local curricular capacity in international development cooperation.*
93. Benedictus, F. (2017). *Reichenbach: Probability & the A Priori. Has the Baby Been Thrown Out with the Bathwater?*
92. Ruiter, Peter de (2016). *Het Mijnwezen in Nederlands-Oost-Indië 1850-1950.*
91. Roersch van der Hoogte, Arjo (2015). *Colonial Agro-Industrialism. Science, industry and the state in the Dutch Golden Alkaloid Age, 1850-1950.*
90. Veldhuis, M. (2015). *Improving classroom assessment in primary mathematics education.*
89. Jupri, Al (2015). *The use of applets to improve Indonesian student performance in algebra.*
88. Wijaya, A. (2015). *Context-based mathematics tasks in Indonesia: Toward better practice and achievement.*
87. Klerk, S. (2015). *Galen reconsidered. Studying drug properties and the foundations of medicine in the Dutch Republic ca. 1550-1700.*
86. Krüger, J. (2014). *Actoren en factoren achter het wiskundecurriculum sinds 1600.*
85. Lijnse, P.L. (2014). *Omzien in verwondering. Een persoonlijke terugblik op 40 jaar werken in de natuurkundedidactiek.* Utrecht University, Utrecht.
84. Weelie, D. van (2014). *Recontextualiseren van het concept biodiversiteit. Utrecht University, Utrecht.*
83. Bakker, M. (2014). *Using mini-games for learning multiplication and division: a longitudinal effect study.*
82. Ngô Vũ Thu Hăng (2014). *Design of a social constructivism-based curriculum for primary science education in Confucian heritage culture.*

81. Sun, Lei (2014). *From rhetoric to practice: enhancing environmental literacy of pupils in China.*
80. Mazereeuw, M. (2013). *The functionality of biological knowledge in the workplace. Integrating school and workplace learning about reproduction.*
79. Dierdorp, A. (2013). *Learning correlation and regression within authentic contexts.*
78. Dolfing, R. (2013). *Teachers' Professional Development in Context-based Chemistry Education. Strategies to Support Teachers in Developing Domain-specific Expertise.*
77. Mil, M.H.W. van (2013). *Learning and teaching the molecular basis of life.*
76. Antwi, V. (2013). *Interactive teaching of mechanics in a Ghanaian university context.*
75. Smit, J. (2013). *Scaffolding language in multilingual mathematics classrooms.*
74. Stolk, M.J. (2013). *Empowering chemistry teachers for context-based education. Towards a framework for design and evaluation of a teacher professional development programme in curriculum innovations.*
73. Agung, S. (2013). *Facilitating professional development of Madrasah chemistry teachers. Analysis of its establishment in the decentralized educational system of Indonesia.*
72. Wierdsma, M. (2012). *Recontextualising cellular respiration.*
71. Peltenburg, M. (2012). *Mathematical potential of special education students.*
70. Moolenbroek, A. van (2012). *Be aware of behaviour. Learning and teaching behavioural biology in secondary education.*
69. Prins, G.T., Vos, M.A.J. & Pilot, A. (2011). *Leerlingpercepties van onderzoek & ontwerpen in het technasium.*
68. Bokhove, Chr. (2011). *Use of ICT for acquiring, practicing and assessing algebraic expertise.*
67. Boerwinkel, D.J. & Waarlo, A.J. (2011). *Genomics education for decision-making. Proceedings of the second invitational workshop on genomics education, 2-3 December 2010.*
66. Kolovou, A. (2011). *Mathematical problem solving in primary school*.
65. Meijer, M. R. (2011). *Macro-meso-micro thinking with structure-property relations for chemistry. An explorative design-based study*.
64. Kortland, J. & Klaassen, C. J. W. M. (2010). *Designing theory-based teaching-learning sequences for science. Proceedings of the symposium in honour of Piet Lijnse at the time of his retirement as professor of Physics Didactics at Utrecht University*.
63. Prins, G. T. (2010).*Teaching and learning of modelling in chemistry education. Authentic practices as contexts for learning*.
62. Boerwinkel, D. J. & Waarlo, A. J. (2010). *Rethinking science curricula in the genomics era. Proceedings of an invitational workshop*.
61. Ormel, B. J. B. (2010). *Het natuurwetenschappelijk modelleren van dynamische systemen. Naar een didactiek voor het voortgezet onderwijs*.

60. Hammann, M., Waarlo, A. J., & Boersma, K. Th. (Eds.) (2010). *The nature of research in biological education: Old and new perspectives on theoretical and methodological issues – A selection of papers presented at the VIIth Conference of European Researchers in Didactics of Biology*.
59. Van Nes, F. (2009). *Young children's spatial structuring ability and emerging number sense*.
58. Engelbarts, M. (2009*). Op weg naar een didactiek voor natuurkunde-experimenten op afstand. Ontwerp en evaluatie van een via internet uitvoerbaar experiment voor leerlingen uit het voortgezet onderwijs*.
57. Buijs, K. (2008). *Leren vermenigvuldigen met meercijferige getallen*.
56. Westra, R. H. V. (2008). *Learning and teaching ecosystem behaviour in secondary education: Systems thinking and modelling in authentic practices*.
55. Hovinga, D. (2007). *Ont-dekken en toe-dekken: Leren over de veelvormige relatie van mensen met natuur in NME-leertrajecten duurzame ontwikkeling*.
54. Westra, A. S. (2006). *A new approach to teaching and learning mechanics*.
53. Van Berkel, B. (2005). *The structure of school chemistry: A quest for conditions for escape*.
52. Westbroek, H. B. (2005). *Characteristics of meaningful chemistry education: The case of water quality*.
51. Doorman, L. M. (2005). *Modelling motion: from trace graphs to instantaneous change*.
50. Bakker, A. (2004). *Design research in statistics education: on symbolizing and computer tools*.
49. Verhoeff, R. P. (2003). *Towards systems thinking in cell biology education*.
48. Drijvers, P. (2003). *Learning algebra in a computer algebra environment. Design research on the understanding of the concept of parameter*.
47. Van den Boer, C. (2003). *Een zoektocht naar verklaringen voor achterblijvende prestaties van allochtone leerlingen in het wiskundeonderwijs*.
46. Boerwinkel, D.J. (2003). *Het vormfunctieperspectief als leerdoel van natuuronderwijs. Leren kijken door de ontwerpersbril.*
45. Keijzer, R. (2003). *Teaching formal mathematics in primary education. Fraction learning as mathematising process*.
44. Smits, Th. J. M. (2003). *Werken aan kwaliteitsverbetering van leerlingonderzoek: Een studie naar de ontwikkeling en het resultaat van een scholing voor docenten*.
43. Knippels, M. C. P. J. (2002). *Coping with the abstract and complex nature of genetics in biology education – The yo-yo learning and teaching strategy*.
42. Dressler, M. (2002). *Education in Israel on collaborative management of shared water resources.*
41. Van Amerom, B.A. (2002). *Reinvention of early algebra: Developmental research on the transition from arithmetic to algebra*.
40. Van Groenestijn, M. (2002). *A gateway to numeracy. A study of numeracy in adult basic education.*

39. Menne, J. J. M. (2001). *Met sprongen vooruit: een productief oefenprogramma voor zwakke rekenaars in het getallengebied tot 100 – een onderwijsexperiment*.
38. De Jong, O., Savelsbergh, E.R. & Alblas, A. (2001). *Teaching for scientific literacy: context, competency, and curriculum.*
37. Kortland, J. (2001). *A problem-posing approach to teaching decision making about the waste issue.*
36. Lijmbach, S., Broens, M., & Hovinga, D. (2000). *Duurzaamheid als leergebied; conceptuele analyse en educatieve uitwerking*.
35. Margadant-van Arcken, M. & Van den Berg, C. (2000). *Natuur in pluralistisch perspectief – Theoretisch kader en voorbeeldlesmateriaal voor het omgaan met een veelheid aan natuurbeelden*.
34. Janssen, F. J. J. M. (1999). *Ontwerpend leren in het biologieonderwijs. Uitgewerkt en beproefd voor immunologie in het voortgezet onderwijs*.
33. De Moor, E. W. A. (1999). *Van vormleer naar realistische meetkunde – Een historisch-didactisch onderzoek van het meetkundeonderwijs aan kinderen van vier tot veertien jaar in Nederland gedurende de negentiende en twintigste eeuw*.
32. Van den Heuvel-Panhuizen, M. & Vermeer, H. J. (1999). *Verschillen tussen meisjes en jongens bij het vak rekenen-wiskunde op de basisschool – Eindrapport MOOJ-onderzoek*.
31. Beeftink, C. (2000). *Met het oog op integratie – Een studie over integratie van leerstof uit de natuurwetenschappelijke vakken in de tweede fase van het voortgezet onderwijs*.
30. Vollebregt, M. J. (1998). *A problem posing approach to teaching an initial particle model*.
29. Klein, A. S. (1998). *Flexibilization of mental arithmeticsstrategies on a different knowledge base – The empty number line in a realistic versus gradual program design*.
28. Genseberger, R. (1997). *Interessegeoriënteerd natuur- en scheikundeonderwijs – Een studie naar onderwijsontwikkeling op de Open Schoolgemeenschap Bijlmer*.
27. Kaper, W. H. (1997). *Thermodynamica leren onderwijzen*.
26. Gravemeijer, K. (1997). *The role of context and models in the development of mathematical strategies and procedures*.
25. Acampo, J. J. C. (1997). *Teaching electrochemical cells – A study on teachers' conceptions and teaching problems in secondary education*.
24. Reygel, P. C. F. (1997). *Het thema 'reproductie' in het schoolvak biologie*.
23. Roebertsen, H. (1996). *Integratie en toepassing van biologische kennis – Ontwikkeling en onderzoek van een curriculum rond het thema 'Lichaamsprocessen en Vergift'*.
22. Lijnse, P. L. & Wubbels, T. (1996). *Over natuurkundedidactiek, curriculumontwikkeling en lerarenopleiding*.

21. Buddingh', J. (1997). *Regulatie en homeostase als onderwijsthema: een biologie-didactisch onderzoek*.
20. Van Hoeve-Brouwer G. M. (1996). *Teaching structures in chemistry – An educational structure for chemical bonding*.
19. Van den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education*.
18. Klaassen, C. W. J. M. (1995). *A problem-posing approach to teaching the topic of radioactivity*.
17. De Jong, O., Van Roon, P. H. & De Vos, W. (1995). *Perspectives on research in chemical education.*
16. Van Keulen, H. (1995). *Making sense – Simulation-of-research in organic chemistry education*.
15. Doorman, L. M., Drijvers, P. & Kindt, M. (1994). *De grafische rekenmachine in het wiskundeonderwijs*.
14. Gravemeijer, K. (1994). *Realistic mathematics education*.
13. Lijnse, P. L. (Ed.) (1993). *European research in science education*.
12. Zuidema, J. & Van der Gaag, L. (1993). *De volgende opgave van de computer*.
11. Gravemeijer, K, Van den Heuvel Panhuizen, M., Van Donselaar, G., Ruesink, N., Streefland, L., Vermeulen, W., Te Woerd, E., & Van der Ploeg, D. (1993). *Methoden in het reken-wiskundeonderwijs, een rijke context voor vergelijkend onderzoek*.
10. Van der Valk, A. E. (1992). *Ontwikkeling in Energieonderwijs*.
9. Streefland, L. (Ed.) (1991). *Realistic mathematics education in primary schools*.
8. Van Galen, F., Dolk, M., Feijs, E., & Jonker, V. (1991). *Interactieve video in de nascholing reken-wiskunde.*
7. Elzenga, H. E. (1991). *Kwaliteit van kwantiteit*.
6. Lijnse, P. L., Licht, P., De Vos, W. & Waarlo, A. J. (Eds.) (1990). *Relating macroscopic phenomena to microscopic particles: a central problem in secondary science education*.
5. Van Driel, J. H. (1990). *Betrokken bij evenwicht*.
4. Vogelezang, M. J. (1990). *Een onverdeelbare eenheid*.
3. Wierstra, R. F. A. (1990). *Natuurkunde-onderwijs tussen leefwereld en vakstructuur*.
2. Eijkelhof, H. M. C. (1990). *Radiation and risk in physics education*.
1. Lijnse, P. L. & De Vos, W. (Eds.) (1990). *Didactiek in perspectief*.