Ethics in action: Approving and improving medical research with human subjects

de Jong, Jean Philippe

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
In this thesis, Jean Philippe de Jong presents a new understanding of ethical oversight on medical research with human subjects and proposes that there are two philosophies for ethical oversight: ‘(dis)approving’ and ‘improving’. Systems for ethical oversight on medical research have been in place for many years, with Research Ethics Committees as their cornerstone. Although these oversight systems aim to ensure that the ethical quality of research is in order, they have been criticized for impeding the progress of science and failing to provide adequate protection for research subjects. In reality however, little is known about the practice of ethical oversight: it is a ‘black box’. This thesis opens up this black box and studies different aspects of ethical oversight in practice. It deals, amongst others, with ethical review of research proposals, oversight on the actual conduct of research, (non-)publication of research and justice in distributing the benefits of research results. The author suggests there are two philosophies for doing good in ethical oversight: (1) ‘(dis)approving’ which aims to ensure that research is ethical and works towards (dis)approval, is carried out within relationships based on authority and focuses on documentation; and (2) ‘improving’ which aims to improve the ethical quality of research by giving advice, is carried out within relationships based on equality and focuses on the actual research practice. The author concludes that although (dis)approving is closer to what people expect from Research Ethics Committees and is the dominant approach to ethical oversight, improving is a valuable alternative for supporting scientific progress and guarding the interests of research subjects.
ETHICS IN ACTION

Approving and Improving Medical Research with Human Subjects

JEAN PHILIPPE DE JONG
ETHICS IN ACTION
Approving and Improving Medical Research with Human Subjects

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college
voor promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 11 december 2013, te 14:00 uur

door

Jean Philippe de Jong

geboren te Rotterdam
PROMOTIECOMMISSIE

Promotor
Prof. dr. D.L. Willems

Overige leden
Prof. dr. J.J.M. Van Delden
Prof. mr. J.C.J. Dute
Prof. dr. D.A. Legemate
Prof. dr. M.M. Levi
Prof. dr. M.H.N. Schermer
Prof. dr. J.G.P. Tijssen

Faculteit der Geneeskunde
**TABLE OF CONTENTS**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prologue</td>
<td>7</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>9</td>
</tr>
<tr>
<td>The black box of ethical oversight on medical research and finding a way in</td>
<td></td>
</tr>
<tr>
<td>2 Ethical Review from the Inside</td>
<td>25</td>
</tr>
<tr>
<td>Repertoires of evaluation in Research Ethics Committee meetings</td>
<td></td>
</tr>
<tr>
<td>3 Compliance or Quality Improvement?</td>
<td>49</td>
</tr>
<tr>
<td>Research monitoring by US medical institutions to protect human subjects</td>
<td></td>
</tr>
<tr>
<td>4 Publication Rate of Clinical Studies</td>
<td>69</td>
</tr>
<tr>
<td>Two prognostic indicators of the publication rate of clinical studies were available during ethical review</td>
<td></td>
</tr>
<tr>
<td>5 Tumour Tissue: Who is in Control?</td>
<td>91</td>
</tr>
<tr>
<td>Guidelines on tissue banking for clinical and research purposes</td>
<td></td>
</tr>
<tr>
<td>6 Justice in Research Methods</td>
<td>103</td>
</tr>
<tr>
<td>Standardization in randomized controlled trials</td>
<td></td>
</tr>
<tr>
<td>7 Discussion</td>
<td>125</td>
</tr>
<tr>
<td>(Dis)approving and improving in action</td>
<td></td>
</tr>
<tr>
<td>Epilogue</td>
<td>145</td>
</tr>
<tr>
<td>Supplement</td>
<td>147</td>
</tr>
<tr>
<td>Summary</td>
<td>163</td>
</tr>
<tr>
<td>Samenvatting</td>
<td>169</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>175</td>
</tr>
<tr>
<td>Portfolio</td>
<td>179</td>
</tr>
<tr>
<td>List of Publications</td>
<td>183</td>
</tr>
<tr>
<td>Curriculum Vitae</td>
<td>185</td>
</tr>
</tbody>
</table>
An elderly man, for confidentiality reasons I will call him Jan, went to see his general practitioner with mild urinary tract symptoms. Jan was sent for a biopsy, which confirmed he had prostate cancer. After consulting a surgeon, Jan decided that he wanted to have 'nerve-sparing' surgery, mainly because it provided a better chance of preserving his erectile function compared to a more radical removal.

However, the surgeon also asked Jan whether he would be willing to participate in a scientific study to collect and store tissue and patient data for use in future research. In addition to storing blood samples and tumor tissue, the study involved additional imaging of the prostate with MRI, a diagnostic technique that was not part of regular clinical care for prostate cancer. The main risks of participating in this study were the potential side-effects of the contrast agent used for the MRI. The informed consent form also mentioned that the results of the MRI could, in individual cases, be used to adapt the therapeutic approach and the choice for a particular surgical procedure. After reading the consent form, Jan decided to participate since he thought the risks were small and he wanted to help improve the treatment of future patients.

Unfortunately, Jan’s MRI indicated that nerve-sparing surgery would not result in a removal of all tumor tissue, so the surgeon performed a more radical operation, including a removal of the nerves. Jan recovered well after the operation, but suffered from permanent impotency. Pathological analysis after the operation showed that the MRI had given misleading information and that nerve-sparing surgery would have been possible after all. The surgeon told Jan that, within the context of regular clinical care, he would indeed have chosen to perform a nerve-sparing procedure. So, in effect, participating in the scientific study had been a major causal factor for Jan’s impotency.
Although Jan appreciated that unfortunate events happen in scientific research, and he did not feel that the surgeon or the investigative team was culpable for his impotency, he was left with many questions. Had it been a good idea to allow an alteration of his treatment based on the MRI? Was he informed properly about the risks or could additional information have changed his mind? Was the knowledge gained by this study worth the risks to him and other patients? And he also started wondering: Who decides about all of these issues? Who ensures that these things are in order? And how do they do that?
INTRODUCTION

The black box of ethical oversight on medical research and finding a way in

1.1 Ethical oversight on medical research with human subjects

This thesis studies how ethical oversight on medical research with human subjects functions in practice. Fueled by public outrage about experiments by the NAZI’s, the Tuskegee Syphilis Study, Henry Beecher’s revelation of unethical research practices at leading hospitals in the United States, and other such scandals, the need for independent oversight on the ethical quality of research with humans has long been highlighted in leading ethical guidance documents, e.g. the Declaration of Helsinki and the Belmont Report (18th WMA General Assembly 1964, Beecher 1966, Jones 1993, The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979). The ethical quality of medical research depends on whether it is in line with relevant moral principles: the social value of research, the scientific validity, a fair subject selection, a favorable ratio of risks and burdens to subjects and benefits for society, independent ethical review, informed consent of subjects, and respect for enrolled subjects (Emanuel et al. 2000). I define ethical oversight as the activities of independent organizations that work on the ethical quality of medical research.

In The Netherlands, as in most other countries, Research Ethics Committees (in Dutch called ‘Medisch Ethische Toetsingscommissies’; in the U.S. ‘Institutional Review Boards’) have become the cornerstone of ethical oversight on research. These committees review research proposals for medical research to help protect the rights and welfare of human subjects. Research Ethics Committees should function independently and often have a legal basis (in The Netherlands the Medical Research
Involving Human Subjects Act). In practice, most committees are linked to a research institution. Other organizations also contribute to overseeing the ethical quality of medical research. In The Netherlands, ethical oversight is also performed by the Central Committee on Research involving Human Subjects, the Health Care Inspectorate, public funding bodies, pharmaceutical companies, offices at research institutions and research leaders. Still, Research Ethics Committees are the most visible and authoritative bodies overseeing the ethical quality of research. Moreover, they are the only bodies that independently review all proposals for medical research that could affect the rights and welfare of human subjects.

According to Dutch law, the task of a Research Ethics Committee is, amongst others, to evaluate and judge proposals for medical research, based on considerations about the contribution of the study to scientific progress in relation to the interests of research subjects, including the burdens and risks of participating in the study (article 3, Wet Medisch-wetenschappelijk Onderzoek met Mensen 1998). Researchers can only start conducting a study after a committee has given formal approval. Furthermore, according to the Declaration of Helsinki, Research Ethics Committees also have the right to oversee (‘monitor’) ongoing studies.

Although Research Ethics Committees are generally considered to play an important role in overseeing the ethical quality of research (18th WMA General Assembly 1964), they have also been criticized (Ashcroft and Pfeffer 2001). There has been frustration in the research community, among regulatory authorities, and in the pharmaceutical industry about the efficiency of committees, and some feel that they unnecessarily impede the progress of science (Fost and Levine 2007, Koski 2003, Saunders 2002, Savulescu 2002, Shalala 2000, Steinbrook 2002). At the same time, there have also been doubts about whether Research Ethics Committees provide adequate protection for research subjects (Fost and Levine 2007, Koski 2003, Saunders 2002, Savulescu 2002, Shalala 2000, Steinbrook 2002).
1.2. The black box of ethical oversight

An important factor fueling the discussions about Research Ethics Committees is that little is publicly known about their functioning. Although Research Ethics Committees are to play an important role in safeguarding the interest of subjects in research and thus have a social function, committees take up this responsibility predominantly behind closed doors (Ashcroft and Pfeffer 2001). Many aspects of how Research Ethics Committees should work are prescribed by legislation on European (e.g. the Directive on Good Clinical Practice) and national levels, ethical guidance documents (e.g. the Declaration of Helsinki), and rules and regulations issued by Research Ethics Committees themselves (e.g. regulations governing committee activities, procedures for submitting research proposals, lists of documents to be submitted for review, procedures for lodging appeals and objections). These documents state that committees are to convene and discuss research proposals, specify what criteria committees should use to review proposals, describe procedural aspects of decision making (e.g. voting), and require that the committee notifies researchers of decisions in writing (Centrale Commissie voor Medisch-wetenschappelijk Onderzoek 2003, Code of Federal Regulations 2009, Wet Medisch-wetenschappelijk Onderzoek met Mensen 1998).

But how Research Ethics Committees actually work in practice is not public knowledge. Although authoritative handbooks on the management and function of Research Ethics Committees provide detailed information on the organization of committees and the process of review, they do not elucidate what actually happens during committee meetings, how criteria are applied, how discussions progress, and how decisions are reached (Amdur and Bankert 2002). Furthermore, committee meetings are closed to the public and both minutes of meetings, and letters sent to researchers, remain confidential (Centrale Commissie voor Medisch-wetenschappelijk Onderzoek 2003, Dixon-Woods et al. 2007).

Although I acknowledge that there might be good reasons for keeping meetings (partly) confidential (e.g. information in research proposals can be commercially sensitive and this warrants some sort of protection) and thus make meetings not directly accessible to the public, scientific knowledge of the practice of ethical oversight is also limited (Abbott and Grady 2011, Sheehan 2008). A 2011 systematic review of studies
that evaluated aspects of the functioning of U.S. Institutional Review Boards (IRB) collected forty-three empirical studies, which focused on IRB volume, characteristics of IRB members, costs associated with review, the extent to which federal regulations were implemented, variation in review of multicenter research, and the results of IRB deliberations (Abbott and Grady 2011). Information in these studies mainly came from surveys, interviews, institutional databases, and IRB records and letters. However, within these forty-three empirical studies, investigations of the actual practice of doing research oversight were virtually absent and researchers rarely visited the sites where review takes place, i.e. committee meetings. During the last two decades, the only study that visited IRB sites was a 1998 report by the Department of Health and Human Services (Department of Health and Human Services 1998). This study found that IRBs review too quickly, with little expertise and training, that conflicts threaten their independence, and that minimal continuing review occurs. Outside the U.S., the same picture emerges: empirical studies have been studying ethical oversight ‘from the outside’, leaving the actual practice of doing oversight a black box (De Jong et al. 2011). It thus remains essential to study in practice how Research Ethics Committees accomplish their objectives (Abbott and Grady 2011).

So, because limited public and scientific knowledge is available about how ethical oversight works in practice, we are, just as my neighbor Jan mentioned in the prologue, left wondering how the system of ethical oversight on medical research works. How does this system serve the public good and protect the interest of the – in The Netherlands alone – tens of thousands of persons who participate in research each year? This thesis takes up that question.

The central aim of this thesis is to help understand how ethical oversight, especially by Research Ethics Committees, works. A secondary aim of this thesis is to help improve the functioning of ethical oversight. This thesis therefore takes up the following questions: How do Research Ethics Committees evaluate proposals for medical research during their meetings? Do Research Ethics Committees and research institutions oversee whether the actual conduct of research is ethical, and if so, how? Do approved studies actually lead to scientific knowledge and how can Research Ethics Committees oversee that they do? How to handle cases where patients’ medical interests conflict with the interest of scientific progress? And how can oversight bodies
make sure that the benefits of research are distributed fairly among the population, thus serving the public good? Taken together, answering these questions will help provide insight in how ethical oversight on medical research works in practice.

1.3 Looking for a way into the black box

How to study ethical oversight on medical research? Traditionally, the study of moral questions has been carried out by the philosophical discipline of ethics, considering moral issues in health care and research as something to which moral theories could be ‘applied’. However, this ‘applied ethics’ model has been criticized for failing to appreciate the circumstances and contexts that generate and structure issues, and the sometimes idiosyncratic reality of work in health care (Hoffmaster 1992, Ten Have and Lelie 1998). Furthermore, I believe that morality emerges from the practice of daily life rather than from transcendent theories (Dennett 2003, Ridley 1996), so to understand morality, it has to be studied in practice. What is more, studying morality within a practice allows one to ‘articulate’ – give words to – ideals and their related practices that remained implicit; opening up these ideals and practices for reflection (Pols 2008). In health care these ideas have led to an ‘empirical turn’ in ethics, encouraging ethicist to study phenomena in practice instead of from theory (Willems and Pols 2010). This turn has had its effect on the body of medical ethics literature: in the 1980ies 8% of studies had an empirical approach, rising to 16% at the beginning of this millennium (Sugarman et al. 2010). This thesis’ quest to understand how ethical oversight on medical research works in practice follows this empirical turn.

So, how should we enter the ‘black box’ of ethical oversight? And when we enter, where should we look in order to understand what is going on inside? For studying the practice of ethical oversight, I was inspired by the methods used in the field of Science and Technology Studies (STS), in particular the ones described by one of the founders of the field, Bruno Latour, in ‘Science in Action’ (Latour 1987). In this book Latour demonstrates how social context and technological content are essential to a proper understanding of scientific facts and their production. He emphasizes that the production of scientific facts can only be understood by studying science as a practice
and examines science *in action*: the activities in research laboratories, the role of scientific literature, the institutional context of science in the modern world, and the means by which inventions and discoveries become accepted. To paraphrase this in terms of ethical oversight: ethical oversight can only be understood by studying it as a practice, by examining oversight *in action*: the activities of Research Ethics Committees, the role of ethical texts, the institutional context of ethical oversight, and the means by which research becomes ethically acceptable.

Latour’s emphasis on practice suited the aim of this thesis and encouraged me to follow his first rule of method: “We study science *in action* and not ready made science or technology; to do so, we either arrive before the facts and machines are black boxed or we follow the controversies that reopen them.” Latour used controversies as a way to get inside the practice of science. He looked for contrasting positions that turn into tensions between scientists, leading to open debate, visible to outsiders. I have taken up this lead and have looked for the *contrasts* – oppositions that emphasize differences – within the practice of ethical oversight, in order to describe different ‘ways of doing good’. The hypothesis that there were contrasts present in ethical oversight was also inspired by the idea of ‘pluralism’ in ethics: the idea that there are several, conflicting values which may be equally correct, i.e. that there is no objective way to order them in terms of importance (Berlin 1969, Walzer 1983). I have deviated from Latour’s approach since I did not focus on explicit controversies about ethical issues in research, but, keeping in line with the ideas of empirical ethics, have sought to articulate the more implicit contrasts. An additional motivation for focusing on implicit contrast was that public controversies regarding the ethics of research are few and far between. Although public controversies (e.g. the ones mentioned in the first paragraph) have had a major influence on how ethical oversight has been arranged (Emanuel and Grady 2007), the resulting institutionalized form of ethical oversight is performed behind closed doors, making controversies invisible. Moreover, differences of opinion are actively avoided in the current oversight system. For example, multicenter studies are to be reviewed in full by only one committee, aiming to avoid conflicts in review. Also, within committees, members strive for unity in decision making (De Jong, 2011, Directive 2001/20/EC ).
Thus, in order to understand how ethical oversight on medical research works, in each of the next five chapters I study different aspects of ethical oversight in practice and analyze the contrasts within them. In some chapters this will be done quite explicitly, and contrasts will be presented as a main result, whereas in other chapters this will be done in a more subtle way.

In the concluding chapter of this thesis I take my analysis of ethical oversight one step further by again following the lead from empirical ethics. Instead of applying moral theory to practice, as ethical studies have classically done, empirical ethics works the other way around: from moral practice to theory. This move is similar to how normativity is approached in Science and Technology Studies. In that field, the study of normativity is considered neither a matter of description arrived at by scientific study, nor of prescription by philosophy (including ethics), but as a matter of ‘re-prescription’: developing new words, stories and theories for what is good (Harbers 2005, Mol et al. 2010, Willems and Pols 2010). In the concluding chapter I take up this lead and synthesize the contrasts I uncovered in the preceding chapters into a more general understanding of ethical oversight.

1.4 Data and Methods

This thesis contains five studies of ethical oversight, each conducted with different methods and based on different data. All five studies have been written as independent journal articles and contain a detailed description of the data and methods used. Therefore, I here describe the data and methods of this thesis on a general level. I also provide some additional background information on the research setting and my reasons for using particular data and methods.

The data for this thesis were collected through work on two projects. Firstly, in the period 2005-2012, I worked on a PhD project on oversight by Dutch Research Ethics Committees, sponsored by the Academic Medical Center in Amsterdam. Dutch data were complemented by a visiting scholarship at New York University in 2008, in order to study oversight in the United States. The Van Walree Fund sponsored this part of the study. Secondly, in 2006-2007, I worked on a project sponsored by the
Netherlands Cancer Institute to develop a guideline for the storage and use of patients’ residual (tumor) tissue. I was responsible for the ethical considerations underlying the guideline. Besides the issue of how to handle cases where patients’ medical interests conflict with the interests of scientific research, the guideline addressed various other issues concerning storage and use of human tissue. The guideline is attached as a supplement to this thesis.

The majority of the data that were collected in The Netherlands were retrieved from an undisclosed large academic medical center (to which I sometimes refer as the West Holland Medical Center), its corresponding Research Ethics Committee, and from the Netherlands Cancer Institute. Data collected in the United States were mainly retrieved from eleven research institutions in the northeast. Data used in this thesis consisted of: scientific literature, documentation (research protocols and submission forms for ethical review; committee websites and annual reports; internal regulations and procedures of committees; committee archives of minutes and correspondence; databases of research proposals; regulations and guidance documents), and ethnographic materials (stakeholder discussion sessions; observations of committee meetings; questionnaires to researchers; interviews with committee members, institutional officials, and scholars).

In order to analyze the data, my methods covered a broad spectrum, ranging from quantitative methods (e.g. a multivariate Cox regression analysis in chapter four) to qualitative methods (e.g. ethnographic techniques in chapter two). I chose to engage with a variety of data and methods because such a ‘multi-method’ approach can help to get a rich understanding of phenomena (Sulmacy and Sugarman 2010). Moreover, it allows for ‘triangulation’ (comparing the outcomes of different methods) to achieve more robust knowledge (Borkan et al. 2007).

1.5 Outline of this thesis

In this introductory chapter I have set the scene of this thesis: the limited knowledge about ethical oversight on medical research with human subjects. I have also described the aim of this thesis – understanding and improving how ethical oversight works in
practice—and my general approach. In the next five chapters I will take up five research questions to study ethical oversight. Together, these questions will help to achieve a better understanding of ethical oversight.

In chapter two I delve into one of the most important, but at the same time least transparent elements of the system of ethical oversight: the deliberations that take place during Research Ethics Committee meetings. I ask how Research Ethics Committees evaluate the proposals for medical research in practice, during their meetings. Finding an answer to this question could help improve how Research Ethics Committees protect the interests of both subjects and science. By sitting in on committee meetings and analyzing the discussions I discovered that committees are involved in two repertoires of evaluation: a repertoire that focuses on rules and judgments, and a repertoire that focuses on knowledge production and advice. I suggest that although the former repertoire is closer to what many expect from Research Ethics Committees, using the two repertoires in conjunction is worth the while, because it helps researchers to improve the ethical quality of research proposals.

In chapter three I move beyond the review of research proposals, and study what ethical oversight on the actual conduct of research amounts to in practice. I ask how Research Ethics Committees and research institutions monitor the conduct of research and why they have arranged it this way. Getting insight in monitoring practices is important because monitoring of ongoing research has been proposed as an additional way of improving the protection of the rights and welfare of research subjects. I studied Research Ethics Committees and research institutions in the U.S. because independent monitoring programs have been in place there for several years, unlike The Netherlands. My analysis showed that monitoring programs varied considerably, but gravitated towards two general types: compliance monitoring, which focuses on documentation, and can amount to disciplining researchers and requiring mandatory corrective actions; and quality improvement monitoring, which focuses more on actual research conduct, and can result in feedback to both researchers and the research institution on how to improve the research process. I argue that quality improvement monitoring is the better choice because it helps foster trust between researchers and Research Ethics Committees, leading to a better protection of the interests of research subjects.
In chapter four I further narrow down the question of how ethical oversight can influence the conduct of research. I investigate to what extent research studies lead to scientific progress, i.e. publications, and how Research Ethics Committees could ensure that they do. I ask to what extent study results are published, and whether a committee could predict failure to publish already during ethical review. Failure to publish is a grave way of treading research subjects’ interests, since failure to publish makes research subjects’ efforts go in vain and can bias the scientific literature. Being able to predict failure to publish could give Research Ethics Committees an important tool with which to better do justice to research subjects. I found that almost half of the studies that had actually included research subjects remain unpublished. Furthermore, by comparing studies that had been published to those that were not, I found that studies that had a problematic review process and studies that aimed to benefit patients directly (as opposed to fundamental research) were associated with publication failure. Research Ethics Committees could use this information to monitor whether studies lead to publication, and for discussing their worries with researchers during the review process in order to prevent non-publication.

In chapter five I study oversight on the use and storage of human tissue and describe the development of a guideline for the management of patients’ residual (tumor) tissue. Residual tissue is often stored for research purposes, but can also sometimes serve clinical ends at a later moment. This can lead to a conflict of interests between patients’ and research interests. This came to the fore when a woman, previously treated for breast cancer, requested her physician to have a new genetic test performed on her residual tumor tissue which had been stored for research. As guidance was lacking for how to handle this case and how to weigh the interests of a patient against those of research, developing a guideline seemed appropriate. The analysis conducted in this study showed that such a guideline should take four ethical principles into account: the responsibility of health care providers to provide good clinical care; the rights of patients regarding their bodily material, removed or not; the relative rights of family members regarding this material; and the overriding interest of patients’ medical interests over the interests of scientific research in cases where they conflict. The practical implications for the management of human tissue were also explored, including the practicalities of storing sufficient tissue for future clinical usage.
In chapter six I examine how considerations of justice can be relevant to ethical oversight and what implications this holds for research. I explore the idea that the choice for a specific research methodology can affect whether the benefits of research results are distributed fairly. Using the case of hypertension management as an example, I show how the current ‘gold-standard’ of research methods – the randomized controlled trial – aims at a standardization of patients, a standardization of care and a standardization of interventions. I argue that this philosophy of ‘standardization’ can mean that less relevant medical knowledge will become available for disadvantaged patient groups. I further argue that less standardization in RCTs could be beneficial, and that research methods that do not tend to standardize, such as qualitative methods, are needed to generate relevant knowledge for disadvantaged groups. These are considerations that could be taken into account in ethical oversight.

In the concluding chapter, chapter seven, I work towards a deeper understanding of ethical oversight, based on a synthesis of the contrasts uncovered in the studies in this thesis. I argue that there are two philosophies for doing good in ethical oversight: (1) ‘(dis)approving’ which aims to ensure that research is ethical and works towards (dis)approval, is carried out within relationships based on authority, and focuses on documentation; and (2) ‘improving’ which aims to improve the ethical quality of research by giving advice, is carried out within relationships based on equality, and focuses on practice. After reflecting on the research methods used in this thesis I discuss the strengths and weaknesses of both approaches. Although (dis)approving currently is the dominant approach in ethical oversight, I argue that improving is a good alternative. Subsequently, I discuss how the two approaches to oversight interact, and how they can best be combined. In conclusion, I articulate the main lessons that can be drawn from this thesis. Ethical oversight can benefit from focusing more on actual research practices and less on paperwork, from avoiding relationships based on authority (if possible) and trying to work on an equal footing with researchers instead, and from improving the ethical quality of research instead of only (dis)approving research.
References

18th WMA General Assembly (1964) World Medical Association Declaration of Helsinki; Ethical Principles for Medical Research Involving Human Subjects.


Savulescu, J. (2002) Two deaths and two lessons: is it time to review the structure and function of research ethics committees?, *Journal of Medical Ethics*, 28, 1, 1-2.


Abstract

Evaluating the practice of ethical review by Research Ethics Committees (RECs) could help protect the interests of human subjects and promote scientific progress. To facilitate such evaluations, we conducted an ethnographic study of how an REC reviews research proposals during its meetings. We observed thirteen meetings of a Dutch REC, and studied REC documents. We coded this material inductively, and categorised these codes in two repertoires of evaluation: a repertoire of rules and a repertoire of production. In the repertoire of rules, the REC applies rules, weighs scientific value and burdens to subjects, and makes a final judgment on a research proposal in a meeting. In the repertoire of production, REC members check documents and forms, give researchers advice on how to improve their proposals and can use informal communication. Based on these findings, we think that evaluations of the practice of ethical review should take into account that RECs can use a repertoire of rules and a repertoire of production to evaluate research proposals. Combining these two repertoires can be a viable option: the REC giving researchers advice on how to improve their proposals to prevent rejection of valuable research.

1 This chapter was published as: De Jong, J.P., Van Zwieten, M.C.B. & Willems D.L. (2011) Ethical review from the inside: repertoires of evaluation in Research Ethics Committee meetings, Sociology of Health & Illness, 34, 7, 1039-1052.
2.1 Introduction

Research ethics committees (RECs) review proposals for medical research involving human subjects in order to deal with the central tension in medical research: protecting the interests of human subjects versus allowing scientific progress. However, the quality of REC review has been criticised from two sides: one side believes that RECs fail to provide adequate protection for research subjects, and the other that RECs unnecessarily impede the research enterprise (18th WMA General Assembly 1964, Fost and Levine 2007, Koski 2003, Saunders 2002, Savulescu 2002, Shalala 2000, Steinbrook 2002). To deal with these criticisms, many authors and governmental organizations have called for an evaluation of REC review (Centrale Commissie Mensgebonden Onderzoek (CCMO) and Central Committee on Research Involving Human Subjects 2009, Coleman and Bouesseau 2008, Department of Health and Human Services 1998, Feldman and Rebholz 2009, Grady 2010, Taylor 2007). To date, the empirical basis for such evaluations remains small. Most previous studies have focused on administrative procedures and efficiency, variation in REC reviews of multicenter research, differences in assessing specific issues or implementing regulations, and on the function of REC letters (Abbott and Grady 2011, Angell et al. 2006, Angell and Dixon-Woods 2009, Dixon-Woods et al. 2007, Dixon-Woods and Angell 2009, Edwards et al. 2007, O’Reilly et al. 2009, Redshaw et al. 1996, Taylor 2007). However, little information is available on an important part of the review process: the evaluation of research proposals during REC meetings (Abbott and Grady 2011). So far, only Parker et al. (2005) and Fitzgerald et al. (2006) have observed how RECs evaluate proposals for medical research and Hedgecoe (2008) has observed review of social science research. This is especially problematic since REC meetings is the place where the tension between protecting subjects and allowing scientific progress is likely to come to the fore. So, to contribute to the empirical basis for determining how to evaluate REC review, we studied how an REC evaluates research proposals during its meetings.
2.2 Background of ethical review by the West Holland REC

To ensure confidentiality, we refer to the REC we studied as the “West Holland REC” and to the corresponding medical centre as the “West Holland Medical Centre.” In the Netherlands, the Medical Research Involving Human Subjects Act (MRA), the European Directive on Good Clinical Practice, the Declaration of Helsinki, and requirements issued by the Central Committee on Research Involving Human Subjects (CCMO) form the regulatory framework for RECs. Dutch RECs are formally independent governmental bodies. However, of twenty-seven Dutch RECs twenty-two, including the West Holland REC, get their funding, facilities, personnel, and workload from one or a few non-profit institutions, and are thus strongly linked to those institutions. The West Holland REC consists of a legal scholar, a methodologist, an ethicist, a pharmacologist, a patient representative, a pharmacist, five physicians (a paediatrician, a neurologist, a surgeon, and two internists), and a nurse. The West Holland REC reviews a wide range of studies, from large multicenter phase III drug trials, through investigator-initiated surgical trials and public health research to qualitative studies. This is comparable to most other Dutch RECs (De Jong 2010). The task of a Dutch REC is to systematically evaluate a research proposal and to approve or reject it. The West Holland REC has internal regulations and procedures for managing the review process and researchers are given specific instructions on how to submit the documentation for their research proposal (the proposal itself, the informed consent form (ICF), and several forms for administrative purposes). To allow members to prepare for meetings, they receive copies of submitted documents one week prior to the scheduled meeting.

Meetings are held twice a month, take on average three hours, are supervised by a chair, and proceed according to a standard agenda: opening, incoming mail, continued review of proposals, review of new proposals, review of amendments to approved studies, advice on feasibility of multicenter studies approved by other Dutch RECs, reports on serious adverse events, and reports and queries. The chair of the REC introduces proposals with a summary and then invites members to comment on the proposal itself, the ICF, and other forms. Members ask other members questions to help understand the proposal, or make evaluative comments on a certain aspect of the
proposal and suggest a course of action. Other members can respond to this, leading to discussions. The chair then summarizes the comments and the envisaged course of action. Ten to fifteen new proposals are discussed per meeting, making a total of about 300 a year. It takes on average two meetings to reach a decision to approve or reject a proposal. Rejection rates are about one percent.

2.3 Method

Between 2007 and 2009, we used ethnographic techniques to study how the West Holland REC evaluates research proposals during its meetings. The first author, JP de Jong, observed thirteen REC meetings, and took detailed and extensive field notes of the conversations and interactions between members. After each meeting, he wrote down a preliminary interpretation: a first impression of the meeting, together with salient observations and a preliminary characterization of the evaluation of proposals. Sound recordings were made of nine meetings to check whether conversations were adequately described in our field notes and to extract verbatim quotes. Our field notes were supplemented with documents from the West Holland REC (its annual reports and website) that contained information about review meetings. Under Dutch law it was not mandatory to obtain ethical approval for our study from an REC. We obtained consent of the West Holland REC to observe and record its meetings.

Qualitative analysis of our field notes was facilitated by coding the notes and retrieving segments with MAXqda2. Our analysis was guided by the inductive techniques described by Strauss and Corbin (1990), and consisted of three steps. First, through inductive, open coding and comparison between REC meetings, JP de Jong identified elements pertaining to the evaluation of research proposals, using the preliminary interpretation to aid the development of these codes. Because little information exists on what kind of evaluative practice takes place during REC meetings, the second step in our analysis was to arrange our codes according to what we considered to be open and straightforward questions: Of which action(s) does evaluation consist? How is evaluation carried out? What is the goal of evaluation? Where and when does evaluation take place? Who evaluates? How are the evaluators
and the evaluated related? Although these questions were developed with an inductive approach, we were also inspired by the theoretical framework for analyzing evaluative practices called “situatd judgment”, developed by the sociologists Boltanski and Thévenot (2006). With this framework one can analyse the “operations persons perform when they resort to criticism and collaborate in the pursuit of a justified agreement” (Boltanski and Thévenot 2000). Boltanski and Thévenot developed this framework by connecting philosophical theories of justice to sociological empirical research. The framework of situated judgment therefore allows for describing different worlds of evaluation—consisting of abstract values and the sociological context—and this broad scope of analysis suited our purposes. Moreira (2005) used this framework to describe the evaluative practice of clinical guidelines development meetings, and from him we borrow the term repertoire of evaluation to denote the types of evaluation we uncovered. So we use repertoires in a different sense than discourse analysis, because we have not analysed our material in terms of the social psychology of language, but in terms of the practice of evaluation, consisting of values and the sociological context (Potter and Wetherell 1987). We were also drawn to the framework of situated judgment because it supported our emerging hypothesis that more than one type of evaluation was present in REC meetings. We hypothesised the existence of two types of evaluation with two images in mind: a court of law passing judgment, and a factory improving its product. To identify a specific repertoire of evaluation and to distinguish between repertoires of evaluation, we performed a third analytical step: categorizing codes into groups, together forming a certain repertoire. We categorised inductively by assessing the coherence and consistency of the emerging repertoires, and deductively by comparing them to Boltanski and Thévenot's (2006) worlds that resembled our repertoires: the civic world and the industrial world. To increase consistency of coding and categorisation, samples of data were independently analysed by a second researcher (MCB Van Zwieten) and the emerging repertoires of evaluation were regularly discussed with the project leader (DL Willems) and fellow researchers. Finally, we gave a presentation on the repertoires of evaluation to the West Holland REC as a means for respondent validation.

In the next three sections, we present our analysis of the evaluation of research proposals during REC meetings by describing two repertoires of evaluation: a repertoire
Table 1: Two repertoires of evaluation

<table>
<thead>
<tr>
<th>Evaluative actions</th>
<th>Repertoire of Rules</th>
<th>Repertoire of Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applying rules and</td>
<td>Checking and advising</td>
<td></td>
</tr>
<tr>
<td>weighing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manner of evaluation</td>
<td>As a unity</td>
<td>Efficient</td>
</tr>
<tr>
<td>Goal of evaluation</td>
<td>Final judgments</td>
<td>Production of knowledge</td>
</tr>
<tr>
<td>Place of evaluation</td>
<td>REC meeting</td>
<td>REC meeting and informal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>communication</td>
</tr>
<tr>
<td>Evaluator</td>
<td>REC</td>
<td>REC and researcher</td>
</tr>
<tr>
<td>Relationship between</td>
<td>Hierarchical</td>
<td>Co-workers</td>
</tr>
<tr>
<td>REC and researcher</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

of rules and a repertoire of production, and consequently how these repertoires interact. Our description follows the analytical questions: Of which action(s) does evaluation consist? How is evaluation carried out? What is the goal of evaluation? Where and when does evaluation take place? Who evaluates? How are the evaluators and the evaluated related?

Primary data in the form of quotations are indicated by quotation marks. Additional information to clarify quotations is between brackets. JP de Jong translated the quotations from Dutch. Table 1 summarizes the two repertoires of evaluation.

2.4 The repertoire of rules

In this section, we describe the first repertoire of evaluation in which research proposals are evaluated: the repertoire of rules. When a research proposal is evaluated within the repertoire of rules, one of the REC’s activities is applying rules. This begins when the REC decides whether or not a research proposal falls within the scope of the Medical Research Involving Human Subjects Act (MRA), and whether the REC, in the
words of its members, is “competent” to “judge” the proposal “formally,” for example when a member says: “They [the researchers] don’t do anything extra [compared to standard patient care], so it’s not [does not fall within the scope of the] MRA,” to which another member adds: “One isn’t allowed to say anything right now.” So, if the committee does not consider itself “competent,” it will not review the research proposal at all. If a research proposal does fall within the scope of the MRA, its further evaluation is also based on rules. Committee members can explicitly refer to laws and regulations to substantiate this evaluation, saying things like: “These are just the rules” and “It’s not allowed by law.” The REC may also refer to more general rules as “principles”: “The general principle should be that if it’s possible [to carry out a study only] with adults [subjects], you must do so. The point is that somebody else has to decide on behalf of these children because they aren’t able to.” Besides the MRA and other “external regulations” (like the Directive on Good Clinical Practice), an additional source of rules are the “rules and regulations of the REC” itself. These include a regulation for REC activities, a procedure for submitting research proposals, a list of criteria for the protocol and the ICF, a list of the documents to be submitted, a procedure for lodging an “appeal and objection” and a “complaints regulation.” Finally, the REC also uses verbal agreements as rules, as illustrated by the following quote: “We have all agreed that we would look more carefully at the exemption from insurance [when a study poses minimal risks to subjects].”

Ideally, in the repertoire of rules each member of the committee applies rules in the same manner and comes to the same conclusions, which makes the committee act as a unity. As the REC regulations state: “The committee strives for unity in its decision making.” When committee members agree in their evaluation and envisaged course of action, this agreement is often implicit, with one member making a remark and the others agreeing tacitly by not protesting. Sometimes a member makes this agreement explicit: “We all agree on this.” Furthermore, committee members consider themselves as belonging to a larger unity—the committee—which is demonstrated by the fact that they often use the plural “we” when evaluating research proposals. This is illustrated by the following quotes: “We approve of this,” “We don’t agree [with this aspect of the research proposal],” and “We should take that into consideration.” To protect the unity of the committee it strives for independence and wants to avoid conflicts of interests. Its
rules and regulations therefore stipulate that “if a member of the committee is involved or has an interest in a research proposal that is to be judged, s/he will not take part in the discussions and decision making about it.” This is evident in cases where the chair passes her position on to a fellow committee member during the evaluation of a proposal in which she is involved personally.

However, committee members can also interpret the rules or the facts to which the rules apply differently, leading to differences in evaluation and endangering the committee’s unity. Differences are often solved swiftly by a committee member providing some additional information or by asking the researcher to do so. However, greater threats to the committee’s ideal of acting as a unity—and thus to judging whether a research proposal should be “approved” or “rejected”—can arise when members differ not so much on the interpretation of rules or facts, but on determining the value of the two key aspects of a proposal: the scientific value and the burdens to subjects. This brings us to another activity typical for the repertoire of rules: weighing. The following is an example of this: “We should weigh the importance [of this knowledge] against the burdens [to participants],” to which another member adds: “He [the researcher] doesn’t remove our concern about the ratio between burden and importance.” As a rule, the scientific value should outweigh the burdens to subjects, as illustrated by the following two quotes: “Although this study is quite burdensome, [it’s] also important research” and “Despite being vague research, it carries very little burden.” If the scientific value does not outweigh the burdens to subjects, the committee will reject the proposal. Although weighing is thus a crucial activity in this repertoire of evaluation, it can also lead to conflicts between members and threaten the unity of the committee. For example, when one member attacks the scientific value and another member counter-attacks based on the burdens to subjects: “When they finish their pilot [study], they [will] know just as much as when they started,” followed by another member’s response: “But that’s ... nothing serious, [it’s] not dangerous [to subjects].” The conflict can also develop the other way around, with one member attacking the burdens to subjects, and another member counter-attacking on the scientific value, for example: “I thought it quite burdensome to the control [subjects],” followed by the response: “I think the research question justifies the burdens.” In case a
conflict cannot be solved by discussion the committee’s last resort to restore its unity is voting.

So, in the repertoire of rules, REC activities consist of applying rules, weighing scientific value and burdens to subjects, and judging whether to approve or reject a research proposal. Besides unity, the committee strives in these activities for “finality.” It works towards a final “conclusion” or “decision.” As one member puts it: “We can only approve or reject; we cannot give provisional approval.” However, the committee can also choose to suspend its decision on a proposal in case it wants changes to the proposal or more information. Then, after getting a response from the researcher, the committee will discuss the proposal again in a later meeting under the agenda item “pending projects.” However, the committee considers this course of events less desirable: “We want to look at this [proposal] again, although I don’t like that conclusion.” The endpoint of the REC’s evaluation is its final judgment—an approval or rejection—after which the researcher can proceed as planned. As the REC rules and regulations state: “If the circumstances remain the same, the judgment passed by the committee remains valid.” If a researcher wants to make changes to a study, s/he has to submit an “amendment” which will be judged by the committee.

In the repertoire of rules, evaluation takes place during REC meetings and when REC members prepare for meetings. As the REC regulations put it: “The researcher who will be directly responsible for the study on human subjects is obliged to have the Research Ethics Committee review the intended study prior to … the actual initiation of the study … The committee will discuss the research proposal … during one of its meetings … [and] inform the researcher of its substantiated … judgment in writing … Lawful decisions can only be made during a meeting.”

The preceding description of the repertoire of rules makes it clear who evaluates and who is evaluated: the REC evaluates the researcher. In this repertoire, evaluation takes place within a hierarchical relationship: REC members comment on proposals by saying researchers ”must” clarify, motivate, or change certain aspects of the research protocol or informed consent. Also, a researcher literally has to “submit” a proposal. Other wordings directed at researchers such as “allow,” “accept,” “permit,” “may,” “reject,” “grant exemption,” “reject,” and “approve” also place the researchers in a hierarchical relationship with the REC. The REC also has a hierarchical relationship
with the Central Committee on Research Involving Human Subjects (CCMO). The hierarchical relationship between the REC and the CCMO—the REC in this case having the lower rank—is illustrated by the fact that the REC’s internal regulations state that “the person who submits [a research proposal for review] … can lodge an appeal against a judgment issued by the … [REC with the CCMO].” Furthermore, during meetings REC members sometimes explicitly refer to the CCMO as a source of authority, as is illustrated by the following two quotes: “Would the CCMO approve of this?” and “Can this be directly deduced from the MRA? Not from the text of the act, but [it can be deduced] from other CCMO documents.”

2.5 The repertoire of production

We now turn to the second repertoire of evaluation in which research proposals are evaluated: the repertoire of production. When a research proposal is evaluated within the repertoire of production, one of the REC’s activities is checking. This is illustrated by the abundant use of forms. The REC’s website provides researchers with detailed information on what documents to submit, including several forms as the “General Review and Registration Form” (GRRF), the “West Holland Medical Centre Appendix” (containing institute specific questions), and the Informed Consent Form (ICF). As specified in the CCMO directives, the REC has to “check the completeness and accuracy of the GRRF form.” According to the CCMO, the purpose of this form is “to support the reviewing committee in quickly listing a number of important points … [and] to function as a checklist for the person submitting the research file … The GRRF form remains an aid.” Upon submission, a secretary checks whether all required documents are present, and notifies the researcher promptly in case the submission is incomplete. During the meeting itself, the REC members check whether forms have been filled out correctly, making comments like: “The GRRF form lacks information on the principal researcher and the amount of blood [that will be taken form the subject]” and “Several answers [to questions on the GRRF form] are incorrect: … funding, phase of research.” The REC also checks whether other documents contain the required elements, as
illustrated by the following two quotes: “[In the ICF] the section on privacy is missing” and “[In the protocol] information on the safety of drug … [X] is missing.”

In addition to checking documents, the REC employs another activity in the repertoire of production: giving advice. Advice is a non-obligatory recommendation and can concern practically any aspect of the research proposal. As an example of advice concerning study methodology: “I propose that they either use a … [placebo] treatment or put more emphasis on the hard outcome measures,” to which another member adds: “We’ll indicate that it would benefit the quality of the study if they use a … [placebo] treatment.” The REC can also advise on how to decrease the burdens to patients or to increase the benefits to patients. As an example of decreasing burdens, a member remarks: “What they do with a CT [computed tomography] scan, they could do with an MRI [magnetic resonance imaging, a scan that poses no radiation risks] as well, we should advise that.” As an example of increasing benefits: “It would be better to do a crossover study, because in that case the patients from this one group can get the other treatment after the [initial] treatment [as well], because neither treatment has a hundred percent chance of success.” Furthermore, the REC can advise on how to improve the wording on the ICF, and so goes beyond judging it as inadequate: “On the ICF it says ‘race,’ [that]’s an offensive word and biologically incorrect, [it’s] better [to say] ‘ethnicity’.” Our final examples of giving advice concern efficiency and speed. We will elaborate on these aspects of the repertoire of production in the following paragraphs. Advice in terms of efficiency is illustrated by the following quote: “We’ll advise against using placebos, because in my experience they’re terribly expensive.” And the following is advice in terms of speeding up the research process: “It would be better if they did a randomised study and skipped this step.”

In the repertoire of production, the goals of evaluation are cast in terms of production such as speed and volume. To start with the former, the REC considers the speed of the review process to be an important goal, as is evidenced by an example from its annual report: “[The] advantage of this method of working is that the review process is not prolonged unnecessarily,” and as a member says in a meeting: “That was fast, going through such a big stack.” Also in its annual report, the REC stresses its speed by saying that “the submitters usually receive the committee’s written response one week after the committee meeting,” and by presenting a diagram showing the time span from
submission until final decision for each proposal. Furthermore, under the heading “Looking forward,” the REC presents modifications to “accelerate” the review process. Already in the second sentence of its annual report, the REC proudly announces that it produces the highest volume of reviews: “The REC of the … [West Holland Medical Centre] is still the committee that reviews the highest number of scientific protocols per year [within the Netherlands].” To complete this image of the production of evaluations by the REC, we note that its annual report consists for a large part of figures on which categories of research it reviews, in what numbers, what the “throughput time” is, and how much meeting time the REC spends on reviewing. All of this is expressed in numbers, percentages, averages, and graphics. Furthermore, to increase the speed of review and decrease researchers’ workload, the REC takes several measures to help researchers submit proposals correctly. These measures include providing detailed information on what to submit, when to submit it, how to submit it, what regulations apply, examples of protocols, notifying the researcher promptly in case of an incomplete submission, and allowing digital submission of protocols. Also, according to the annual report: “If the committee has no important questions concerning the content, but does have a number of ‘technical’ remarks, the application will be dealt with by the committee’s secretary, which can save a considerable amount of time for both the applicant and the committee.”

An aspect of evaluation related to production is efficiency: evaluations should be produced efficiently. So, REC members are not only concerned with the outcome of their work—the evaluations—but also with how this relates to the input they deliver themselves. REC members invest valuable resources (time and effort) and want to do so efficiently, as the following quote illustrates: “[The forms] have been filled out correctly. That’s because somebody’s working on that professionally. [It’s] smart to hire such a person, [it] saves us a lot of time.” The concern for efficiency also extends to the work done by researchers: the REC advises researchers on how to organise their work efficiently. For example, when the REC discusses two related research proposals submitted by the same researcher—study X and study Y—a member remarks: “I think [the researcher] should first carry out study [X because] if … [the result] isn’t reproducible, they don’t have to carry out … study [Y].” The REC also makes its concern for efficient investment of resources explicit in the quote discussed in the
paragraph on giving advice: “We’ll advise against using placebos, because in my experience they’re terribly expensive.” On the other hand, the REC can also deliberately refrain from improving, for reasons of efficiency. This is often the case with ICFs, for example, when the layout of the ICF could be improved and a member says: “We shouldn’t overdo it.”

Evaluation in the repertoire of production does not only take place during meetings, but is more dispersed in place and time. Because REC members are (albeit sometimes distant) colleagues of a researcher, they can use alternative ways to act and communicate, outside the routes dictated by regulations. For example, when a member asks how to contact a researcher: “Is it possible to ask this without writing a letter?” to which a member replies: “I’ll call the researcher.” As an example of choosing an alternative course of action, a member says: “I think the physician should provide this information, and not the trial nurse. But that doesn’t belong here, I’ll tell them myself.” Furthermore, in the repertoire of production, the REC uses letters to give researchers advice on how to improve their study. This can result in a reply from a researcher explaining how s/he followed the advice, but researchers can in their turn also evaluate the REC’s advice by envisaging how it would affect their study, and respond to the REC with a letter proposing an alternative course of action. This can lead to a discussion between the REC and a researcher, and an exchange of multiple letters. This makes the researcher and the discussion via letters part of the evaluation, and evaluation a co-production of the REC and a researcher. In fact it is very rare for the REC to send only one letter during the review process, expressing its final judgment. Moreover, in the repertoire of production, neither the REC’s evaluation nor the exchange of letters is the decisive step in the evaluation of a study. The REC acknowledges that it can only do so much to improve research proposals, and that studies face the ultimate test during their execution, for example when a member remarks: “Despite the fact that it’s a vague study, it’s not very burdensome. That’s not our problem, it’s the researcher’s problem, it will surely result in something.”

As described in the previous paragraph, in the repertoire of production not just the REC but researchers contribute to the evaluation of research proposals too. Together with the fact that the REC evaluates by giving advice to researchers on how to improve their research proposals, makes the relationship between REC and researchers
one of co-workers helping each other, as is illustrated by this quote: “We phrase this as advice, we help them on their way.”

2.6 Interaction between the two repertoires of evaluation

There are several ways for the repertoires of rules and production to relate to each other during REC meetings. The simplest way is peaceful coexistence: both repertoires are present but do not interact. This can mean that the discussion of a particular research proposal takes place entirely within one repertoire, or that the discussion of a proposal alternates between two repertoires. When the repertoires of rules and production do interact, this can result in either conflict or in collaboration. Conflict can arise when evaluation according to the two repertoires leads to incompatible courses of action. Examples of such conflicts are situations in which members think that formally reviewing a study according to the repertoire of rules is a waste of their and the researcher’s time, according to the repertoire of production. To illustrate this, we quote a member: “[For this study] patients have to say ‘aah’ three times. Unfortunately, [this study] falls under the MRA from a legal point of view. We can’t say it doesn’t fall under the MRA. It would be nice if they hadn’t submitted it, no one would have known a thing.” The committee consequently decides to formally review the proposal. The outcome of this and most other conflicts between the two repertoires is that the repertoire of rules prevails and proposals are formally reviewed.

We have come across only one type of conflict where the repertoire of production triumphs over the repertoire of rules. This is when REC members think that although a research proposal has been judged too harshly, according to the repertoire of rules, a reversal of the judgment is a waste of energy, according to the repertoire of production. To illustrate this: “According to the law this isn’t drug research … But if they’re going to change this, they’ll have to go to the CCMO, that’s a lot of work. We could also pretend we didn’t notice this. They’ve done more than necessary, there’s no harm in that.” However, this type of conflict can also be interpreted as a collaboration between the two repertoires of evaluation: judging a research proposal in a certain way contributes to efficiency. Another example of such a collaboration is the following
quote: “Does this fall under the MRA? It’s only a questionnaire,” to which another member replies: “Let’s judge … [this proposal], [because] we’ve already done the work [preparing for the meeting] anyway.” Although this type of collaboration occurs, the converse relationship between the repertoire of rules and production is far more common: improving a research proposal to be able to judge it in a certain way. An example of this type of collaborative relationship is when an REC member responds to a member that asks how strict the REC is in rejecting a proposal if the ICF is not in order: “We’ve never done that [rejecting a proposal because of an unacceptable informed consent form], … [the form] always turns out all right eventually.” Furthermore, although we have not performed a quantitative analysis, we have the impression that the REC spends roughly the same time evaluating according to either repertoire. Because the REC spends so much time and energy on giving researchers advice on how to improve proposals according to the repertoire of production, this helps to make proposals more approvable, according to the repertoire of rules. Consequently, while it is very rare that a proposal gets approved in the form the researcher initially submitted it, rejection of a proposal is also very rare.

2.7 Discussion

We have described how the West Holland REC evaluates research proposals according to two repertoires of evaluation: a repertoire of rules and a repertoire of production. In the repertoire of rules, the REC applies rules, weighs scientific value and burdens to subjects, and comes to final judgments on research proposals during its meetings. It does so as a unity, and within a hierarchical relationship with researchers. In the repertoire of production, REC members check documents and forms and give researchers—their co-workers—advice on how to improve their proposal. They do so swiftly and efficiently, and can communicate outside formal meetings and letters. When the repertoires conflict, the repertoire of rules usually prevails. However, repertoires can also collaborate: resorting to the repertoire of rules sometimes helps to do work efficiently, but more often, resorting to the repertoire of production helps to get a proposal approved.
We have tried to secure the validity of our study by using different datasets, by having multiple researchers carry out the analysis and by respondent validation, all of which did not show large discrepancies. A limitation of our study is that, although we had no explicit theoretical concepts in mind, our analysis was possibly influenced from the start by prior knowledge of Boltanski and Thévenot’s (2006) theory of situated judgment and the worlds they describe. Apart from studying the meetings and documents which contained explicit information on meetings, we did not gather additional information on the review process such as interviews with committee members or REC letters, for two reasons. First, we focused our research efforts on the meetings themselves because they seemed the most understudied part of the review process. Second, we judged that an analysis of interviews or letters would not have altered our description of what committee members do when they evaluate research proposals during meetings because we followed Boltanski and Thévenot (2006) in taking evaluative practices at face value and not with a critical sociological or psychological stance. However, it would be worthwhile to study how evaluative practices during REC meetings relate to those in REC letters. Comparing our findings to previous studies on REC letters suggests that in letters predominantly a repertoire of rules is used (Dixon-Woods et al. 2007, O’Reilly et al. 2009).

A second limitation of our study is that we have not thoroughly investigated whether additional repertoires of evaluation were present in our material. A cursory exploration did hint at a third type of evaluation. This type of evaluation seemed to revolve around personal relationships, committee members acting as friends and imagining how it would feel to have a relative participate in the proposed research. Although this type of evaluation did not play such a decisive role in the evaluation of proposals as the other two repertoires, it would be worthwhile to further uncover this repertoire.

We believe that our analysis of the West Holland REC applies to other Dutch RECs because they have a similar constitution, work under the same regulations, and most of them are linked to a medical centre in the same way (Centrale Commissie Mensgebonden Onderzoek (CCMO) 2010). Our finding that a REC employs two repertoires of evaluation might also apply to RECs outside the Netherlands, since ethical review seems very similar in REC meetings around the World (Fitzgerald et al. 2006).
Three previous studies have observed the evaluation of research proposals during regular REC meetings. In contrast to our finding of two specific repertoires of evaluation, Parker et al. (2005) concluded that an Australian REC was permeated by a single common “logic” of evaluation, called “the practical logic of reasonableness” of lay-persons. Although this “reasonableness” was referred to explicitly, it was also taken for granted: it needed no further justification. Using our framework of repertoires of evaluation, this finding might be explained as a solution for the possible conflict between underlying repertoires of evaluation: forestalling the conflict by evaluating with a sufficiently general concept with which everybody can agree.

In the second study, Fitzgerald et al. (2006) have described the review process in five countries, based on interviews, observations and documents. Their narrative approach focused on the abstract linguistic structure of comments of individual reviewers. This makes their approach complementary to ours, since we did not focus on the linguistic structure of evaluative comments, but on the practice of evaluations, consisting of values and the sociological context. Unfortunately, Fitzgerald et al. do not provide enough primary data to allow analysis in terms of our repertoires of evaluation. However, we recognise many of the narratives described by Fitzgerald et al. in our material, which supports Fitzgerald’s conclusion that review seems very similar in REC meetings around the world.

In the third study, Hedgecoe (2008) observed how three UK RECs made decisions on social science research. He found that part of the role of the REC was to be supportive and facilitating, offering advice on how to improve research proposals, thereby allowing science to progress. This role seems very similar to our repertoire of production. However, Hedgecoe does not report explicitly on the other role(s) of the RECs, so we cannot ascertain whether they are similar to our repertoire of rules.

Finally, we compare our results to the work of Boltanski and Thévenot (2006) that inspired our analytical approach. Although they proceeded deductively from classic works of political philosophy to empirical material consisting of handbooks for people working in types of business, and although their material is very different form our observations of REC meetings, Boltanski and Thévenot’s “civic” and “industrial world” correspond very well to, respectively our “repertoire of rules” and “repertoire of production.” The civic world, which Boltanski and Thévenot trace back to Rousseau,
centres, like our repertoire of rules, around the unity of people and is rule-governed. This world consists of moral principles, (patient) rights, and people that strive for autonomy and are to be protected. This is the moral world in which most contemporary bioethical debates, including the debates on research ethics, seem to take place. The industrial world, which Boltanski and Thévenot trace back to Saint-Simon, centres, like our repertoire of production, around progress and is concerned with efficient production. This world consists of hard-working people, working together to improve (healthcare) and striving for effective and efficient processes. We think this moral world is undeservedly overlooked by most bioethicists. Our finding that the West Holland REC is involved in two repertoires of evaluation can be explained by the fact that, in the words of the Declaration of Helsinki, it is “an independent committee … in conformity with the laws and regulations,” and also part of a larger research institution. So, on the one hand the REC is an independent committee in the “civic world,” the place where people use a repertoire of rules. On the other hand, the REC is part of a research industry in the “industrial world,” the place where people use the repertoire of production. A good example of this mixed character of RECs is that the research industry’s wish for quick review has in many countries been incorporated in guidelines for REC review by requiring a decision within sixty days. Based on the above comparison with the literature, we conclude that our finding that an REC reviews according to a repertoire of rules and a repertoire of production possibly applies to other RECs around the world. However, future studies that directly observe RECs should confirm this.

We hope to have made it clear that it would be wrong to conflate the two repertoires of evaluation to the tension between the interests of human subjects and the progress of science: both these values partake in both repertoires. We conclude our article by envisaging how the repertoires of evaluation affect both the interests of subjects and science.

*The repertoire of rules and subjects.* In this repertoire the REC strives to apply rules in a uniform manner, and evaluates whether the scientific value outweighs the burdens to individual human subjects. This helps to protect subjects from the extremes of medical research. However, because rules are rigid and general, they can be inadequate for special situations or for special groups of subjects. Furthermore, the
question whether a research proposal is considered ethically permissible is delegated to the REC, which can make researchers feel less responsible.

*The repertoire of rules and science.* This repertoire has the obvious drawback of slowing down scientific progress due to the efforts involved in meeting all regulatory requirements, or by banning types of research. However, when RECs succeed in applying rules in a uniform manner this also has an advantage for researchers: they can carry out research in a predictable environment. Together with the fact that REC decisions are final, this helps researchers to do research with a necessary degree of independence.

*The repertoire of production and science.* In this repertoire the REC gives advice to researchers that can help them to improve their research proposal. The fact that the REC and researchers are co-workers makes this advice practical, tailor-made and allows for additional forms of communication outside formal meetings and letters. Furthermore, the REC tries to do its work efficiently, which helps the progress of research. However, by interfering too much with the research methodology, the REC can endanger the independence of researchers to pursue their research as they see fit.

*The repertoire of production and subjects.* This repertoire has the obvious drawback of being so focused on efficiency and speed that it treats subjects as mere numbers in a calculation and threatens their interests as individuals. However, the REC’s advice to researchers is not only directed at the research methodology, it also serves the interests of subjects, for example, by helping researchers to minimise the burdens.

Our findings also show that how the interests of subjects and science are interrelated is different in both repertoires. In the repertoire of rules the interests of subjects should always prevail over the progress of science, whereas in the repertoire of production they are values which can be optimized more or less independently. In the light of our finding that the repertoire of rules usually prevails over the repertoire of production, this means that, in our study at least, subject protection is the dominant value in REC review.

So, both repertoires of evaluation have advantages and drawbacks for the interests of human subjects and scientific progress. This brings us back to the question of how to evaluate REC review, and more specifically, how to evaluate an REC that uses
a repertoire of rules and a repertoire of production. First of all, we think that anyone who is concerned about the quality of REC review and wants to evaluate the review process should at least take into account that an REC might be involved in more than one repertoire of evaluation. However, whether RECs should use none, one, or both of the repertoires of evaluation we described is up for discussion. Our results do suggest that a combination of repertoires can be a viable option: use the repertoire of production to give researchers advice on how to improve their research proposals in order to prevent rejections of much valuable research when only the repertoire of rules were used.
References

18th WMA General Assembly (1964) World Medical Association Declaration of Helsinki; Ethical Principles for Medical Research Involving Human Subjects.


Koski, G. (2003) Beyond compliance...is it too much to ask?, *IRB, 25, 5, 5-6.*


Saunders, J. (2002) Research ethics committees--time for change?, *Clinical Medicine, 2, 6, 534-538.*

Savulescu, J. (2002) Two deaths and two lessons: is it time to review the structure and function of research ethics committees?, *Journal of Medical Ethics, 28, 1, 1-2.*


3

COMPLIANCE OR QUALITY IMPROVEMENT?¹
Research monitoring by US medical institutions to protect human subjects
Jean Philippe de Jong, Myra C.B. van Zwieten, Dick L. Willems

Abstract

In recent years, to protect the rights and welfare of human subjects, institutions in the United States have begun to set up programmes to monitor ongoing medical research. These programmes provide routine, on-site oversight, and thus go beyond existing oversight such as investigating suspected misconduct or reviewing paperwork provided by investigators. However, because of a lack of guidelines and evidence, institutions have had little guidance in setting up their programmes. To help institutions make the right choices, we used interviews and document analysis to study how and why eleven United States institutions have set up their monitoring programmes. Although these programmes varied considerably, we were able to distinguish two general types. ‘Compliance’ programmes on the one hand were part of the institutional review board (IRB) office and set up to ensure compliance with regulations. Investigators’ participation was mandatory. Monitors focused on documentation. Investigators could be disciplined, and could be obliged to take corrective actions. ‘Quality-improvement’ programmes on the other hand were part of a separate office. Investigators requested to be monitored. Monitors focused more on actual research conduct. Investigators and other parties received feedback on how to improve the research process. Although both

types of programmes have their drawbacks and advantages, we argue that if institutions want to set up monitoring programmes, quality improvement is the better choice: it can help foster an atmosphere of trust between investigators and IRB, and can help raise the standards for the protection of human subjects.

3.1 Introduction

Many authors have argued that ethical review of medical research proposals is not enough to protect the rights and welfare of human subjects, and that independent oversight on the actual conduct of research is required because it can: protect the rights and welfare of subjects; help educate investigators about ethical research conduct; improve the ethical review process; protect institutions from governmental action; and help maintain public confidence in medical research and research ethics committees (RECs) (Berry 1997, Bortolussi and Nicholson 2002, Christakis 1988, Douglass et al. 1998, Heath 1979, Hoffman 2001, Linden-Laufer 1997, McCusker et al. 2001, Norton and Wilson 2008, Smith et al. 1997, Weijer et al. 1995). Following these authors, we call this type of oversight ‘monitoring’ and define it as independent, routine, on-site oversight for human subjects protection. So, monitors are independent from the investigative team, which contrasts with data and safety-monitoring boards; monitoring is a routine activity, which contrasts with investigations of suspected misconduct (‘for-cause audits’); and monitors go on site and contact the investigative team and the project, which contrasts with reviewing paperwork prepared by investigators for yearly re-approval (Heath 1979).

United States’ federal regulations echo this requirement for monitoring and state: ‘An IRB [institutional review board; the US terminology for RECs] shall conduct continuing review of research ... and shall have authority to observe or have a third party observe the consent process and the research’ (Department of Health and Human Services 1997). As US IRBs are part of institutions, this has been interpreted as granting both the institution and the IRB the authority to monitor (Prentice et al. 2006). In line with this requirement, many US institutions have developed monitoring programmes (VandenBosch and Maio 2011). However, there are no guidelines for the frequency,
scope, or operational methods for monitoring. Furthermore, scientific evidence on US monitoring programmes is limited to a brief 2008 report on the programmes at eleven institutions (Bortolussi and Nicholson 2002, VandenBosch and Maio 2011). Evidence from other countries is also scarce. So, institutions have practically no guidance for setting up monitoring programs, which might result in programmes that fail to protect the rights and welfare of human subjects.

To help institutions make the right choices when considering how to set up their monitoring program, we have investigated how US institutions carry out this type of oversight. In this article we give a detailed, systematic overview of the various ways institutions have set up monitoring programmes, and discuss the considerations underlying these programmes.

3.2 Method

In January and February 2008, JJ interviewed IRB members/chairs and other professionals involved in monitoring programmes at eleven US institutions (Table 1), using a structured interview format (Table 2).

To create a mixed dataset we chose six institutions with a medical school and five smaller, non-academic medical centres. The number of active, IRB-approved studies per institution ranged from 630 to 2,200. For practical reasons we focused on the north-eastern US. We also analysed documents on monitoring programmes available on the institutions' websites. These documents were first searched on the basis of the interviews in 2008, followed by systematic extraction of all relevant documents on 18/19 April 2011. Furthermore, five American experts on research ethics and IRBs (who had also served on IRBs in the past) were interviewed to ensure that our sample of institutions did not miss important aspects of monitoring practices in the US. All interviews were transcribed verbatim.

Following the structure of our interview guide, our material was coded inductively by comparing the material from different institutions, and codes were readjusted to obtain an optimal fit between the codes and the material. (Strauss and
Table 1: Study population: eleven US institutions

- New York University School of Medicine
- Johns Hopkins Medicine
- Columbia University Medical Center
- New York State Psychiatric Institute
- Albert Einstein College of Medicine
- Yale University School of Medicine
- Montefiore Medical Center
- Joan and Sanford I. Weill Medical College of Cornell University
- Partners Healthcare System for Brigham and Women’s Hospital, Faulkner Hospital, and Massachusetts General Hospital
- Saint Vincent Catholic Medical Centers
- Children’s Hospital Boston, Harvard Medical School

Table 2: Structured interview format

At this particular institution:

- What is the position of monitors within the organization?
- What is the objective of monitoring?
- How are studies selected for monitoring?
- How does the process of monitoring work?
- What additional tasks do monitors have?
- What are the effects of monitoring?

of monitoring activities and underlying considerations. Hereafter, individual institutions were used as units of analysis to discern different types of monitoring programs. This was inspired by a distinction made by several interviewees between a ‘compliance’ philosophy and a ‘quality-improvement’ philosophy. We decided to apply this distinction systematically, and named the corresponding types of monitoring ‘compliance monitoring’ and ‘quality-improvement monitoring’.

Following our interview guide, we describe the range of monitoring activities in six results sections. Within each section we report whether results apply to compliance monitoring, quality improvement monitoring, or both.

3.3 Results

3.3.1 What is the position of monitors within the organization?
Because one institution did not monitor at all, our results on actual monitoring activities are based on ten institutions.

Monitors were part of the IRB office in eight institutions, and part of an independent office in two institutions. Monitoring programmes in the latter group clearly followed a quality-improvement philosophy, and an interviewee explained why the monitors were independent from the IRB:

‘We want to … maintain… the confidentiality and trust of the investigator … Investigators talk very openly and freely to [the monitor].’

These quality-improvement monitors were called quality-improvement staff, or simply investigators. Monitors who were part of the IRB office followed more of a compliance philosophy, and were called compliance monitors or auditors. Monitoring was often called an audit. An interviewee explained that a shift from compliance monitoring towards quality-improvement monitoring was reflected by changing names:

‘We’re moving away from the term “audit”. We’re calling it “review”, because it’s much friendlier … “Audit” is very bureaucratic, very governmental … Our goal … is to educate them, not to find what they’ve done wrong.’
3.3.2 What is the objective of monitoring?
According to all interviewees, monitoring should protect the rights and welfare of subjects. However, compliance programmes and quality-improvement programmes translated this into different practical objectives. As one interviewee put it, the objective of compliance monitoring was the ‘enforcement of regulatory compliance’. Furthermore, some compliance programmes were initiated after incidents with human subjects and problems with governmental agencies: ‘In the past we found [that investigators] … did not do well on FDA inspections.’ According to this interviewee, monitoring should also protect the institution from governmental interference and legal action:

‘An institution could be held liable if something goes wrong … [If an investigator is not in compliance] an institution is exposed without any [insurance] coverage.’

Quality-improvement programmes framed the goal of subject protection in terms of quality improvement: ‘[to] continually … improve the research process.’ Furthermore, according to our interviewees, quality-improvement monitoring ‘is a service [to investigators]’ and helps ‘to understand where, as an institution, we may be able to improve.’

One interviewee explained why they had chosen a quality-improvement philosophy rather than a compliance philosophy:

‘The culture around here [is] about … openness … Communication and trust just permeates everything [here] with clinical research … I don’t want to disassemble … [this] culture … [That’s] why [our monitors] don’t … go directly to the IRB [with their findings].’

3.3.3 How are studies selected for monitoring?
Four monitoring programmes selected studies at random, four used a risk-based approach, and two used a combined approach. Table 3 summarizes the criteria for selecting studies in risk-based programmes. Selection methods were not related to the type of program. The number of studies monitored ranged from six to over a hundred per year, and the percentage ranged from a few percent to more than 90 percent.
Table 3: Criteria for selecting studies in risk-based monitoring programmes

<table>
<thead>
<tr>
<th>Study-related risk factors:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Studies for which an employee of the institution holds the Investigational New Drug or</td>
</tr>
<tr>
<td>Investigational Device Exemption (and therefore has to fulfil associated FDA regulations)</td>
</tr>
<tr>
<td>• High safety risks to subjects (for example, ‘research that involves recombinant DNA,</td>
</tr>
<tr>
<td>infectious agents and/or pathogens, biological toxins, or gene transfer or pathogens</td>
</tr>
<tr>
<td>introduced into human participants’)</td>
</tr>
<tr>
<td>• Complicated study protocols</td>
</tr>
<tr>
<td>• Old research projects (for example, ‘more than five years’)</td>
</tr>
<tr>
<td>• Investigator-initiated research with no external sponsors</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Investigator-related risk factors:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Studies in which the principal investigator has recently changed</td>
</tr>
<tr>
<td>• Investigators who have come to the attention of the monitor through ‘prior experience</td>
</tr>
<tr>
<td>with the responsible investigator and research team’</td>
</tr>
<tr>
<td>• Investigators who seem to be making mistakes</td>
</tr>
<tr>
<td>• ‘Gut feelings’</td>
</tr>
<tr>
<td>• Investigators new to the institution</td>
</tr>
<tr>
<td>• Inexperienced investigators</td>
</tr>
<tr>
<td>• Investigators who ‘seem slick’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Report-related risk factors:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Problems with continuing review applications and progress reports</td>
</tr>
<tr>
<td>• Inadequate adverse-event reporting</td>
</tr>
</tbody>
</table>

In both types of programmes, studies were also monitored to help ‘prepare for upcoming [external] audits by the FDA, NIH [National Institutes of Health], [or] sponsors.’ Quality-improvement programmes also monitored at the request of investigators and in one programme this was the standard approach; as one interviewee said, ‘We don’t contact people anymore, they call us up: “Would you please come?”’
3.3.4 How does the process of monitoring work?

All monitors sent investigators a letter to announce a monitoring visit. Some monitors specified what information they would need, allowing investigators to prepare for the visit. In quality-improvement programmes visits were voluntary, as phrased on a website: ‘The selected PI [principal investigator] has the option to defer or decline participation,’ whereas in compliance programmes visits were mandatory.

Monitors usually started visits by meeting briefly with the investigator. Visits took from two hours to two days, depending on the complexity of a study and the type of programme: compliance programmes generally taking longer. An interviewee explained why quality-improvement visits were relatively short:

‘If you go there and live with them for a week, nobody wants you to come …
[Furthermore] I trust … [investigators]. I don’t want … [them] to show me everything … [but] some things [that indicate they] … are in good shape. We don’t want to oppress [them].’

All monitors checked whether study records and documents (Table 4) complied with good clinical practice guidelines and regulatory requirements for the protection of human subjects, and corresponded with the IRB’s records.

Quality-improvement monitors also toured the research facility and talked with the research staff. As one interviewee explained:

‘We’ll ask them: … Where are your barriers? What are the problems? … What’s the informed consent process? … That’s the way we get feedback.’

Interviewees of both types of programmes reported that on rare occasions (if the research presented significant risks or if subjects would probably have difficulty understanding relevant information) the informed consent process would be observed.

Final reports were sent to investigators and, in compliance-monitoring programmes, also to other institutional officials. Quality-improvement reports remained confidential unless monitors found ‘serious and continuing non-compliance’, reportable to the authorities and reason to start a for-cause audit. According to a website, quality-improvement reports were made ‘for quality-improvement and educational purposes’ and contained voluntary recommendations to ‘help … improve identified “problem”’.
**Table 4:** Documents checked by monitors

**Basic regulatory documents:**
- Protocol (original and amended versions)
- Consent forms with different versions
- The research team’s curricula vitae, licenses, and certifications
- Laboratory certifications and normal values
- Staff signature log
- Delegation of responsibility log
- Enrolment log
- Documentation of protocol deviations
- Documentation of data- and safety-monitoring activities
- Documentation and assessment of adverse events
- Correspondence with the study sponsor: safety reports, monitoring reports, adverse-event reporting, correspondence on protocol deviations
- Correspondence with the IRB: submitted study documents, approvals of study documents, reports of adverse events, reports of protocol deviations, reports of unanticipated problems

**Additional regulatory documents for investigational drug or device studies:**
- Investigator’s brochure
- Drug/device accountability log
- Correspondence with the FDA

**Individual subject records:**
- Informed consent forms: version, signatures, dates
- Inclusion/exclusion checklists
- Research data
- Source documentation verifying eligibility and research data

**areas**. This contrasts with compliance reports that contained mandatory corrective actions, as is illustrated by this quote:
The report will include any findings that need to be addressed, specified corrective actions, and a time frame by which the action needs to be completed.

Several monitors of both types would follow up on visits by requesting status reports or conducting a second visit. Again, in quality-improvement programmes this was voluntary, whereas in compliance programmes this could be a *mandatory reassessment at a future date*. Furthermore, in compliance programmes, the IRB could take punitive actions against investigators. IRBs did not have *black-and-white, hard rules* to guide such actions, but took several factors into account, as one compliance monitor explained:

'[Were] the subjects really at risk? … Was anybody harmed? … Was it done wilfully? … Was the PI … acting responsibly … or … just left it up to … people who didn’t know what they were doing? … [Has] that researcher … been in trouble with [the IRB] … before?'

IRBs linked to compliance programmes could take various actions: demand good clinical practice training, issue a warning, (temporarily) stop the study, restrict the investigator’s ability to do research, fire the investigator, or, as is illustrated by this quote:

'If [the investigator] … submit[s] … another research protocol [and has] … a bad history with us … [the IRB is] going to be harder on him.'

Quality-improvement programmes however, did not use punitive action, as one interviewee explained:

'We [the monitors] don’t have the authority to do anything, but that’s why … investigators like us … We make recommendations.'

3.3.5 What additional tasks do monitors have?

All but one monitors (a compliance monitor) undertook activities besides monitoring.

Both types of monitors conducted for-cause audits at the request of the IRB or other institutional officials in order to respond to complaints, reports or suspicions of non-compliance (including monitoring reports), or concerns from governmental agencies.
Both types of monitors also had educational tasks, as one interviewee explained: 'Communicating … information to investigators … is … [the monitors’] number-one priority.' Monitors educated through participation in trainings, presentations and symposiums, providing information through newsletters and websites, providing tools for organizing study documentation and providing assistance with study start-up. Although the kind of educational activities overlapped in compliance and quality-improvement programmes, the latter included more educational activities, and focused less on regulations but more on practical advice on running a well-organized study.

Quality-improvement monitors also used the information collected during visits to give the IRB feedback on the review process and institutional officials ‘feedback … about what burdens, what barriers [investigators] have in doing … research’ in order to improve institutional policies and research facilities. An interviewee explained:

‘We look at it [as] … quality-improvement of the whole, any part of the institution that has anything to do with human research protection. When we review an investigator, at the same time we also review the IRB … [their] record [and] … decision.’

Furthermore, feedback was used to improve the quality-improvement monitoring programmes, as another monitor explained:

‘We always … collect information that improves our tools, [to make them] not only efficient for us to use, but also user-friendly for study sites.’

3.3.6 What are the effects of monitoring?
In general, interviewees of both types of programmes reported that monitors seldom found problems concerning subjects’ rights and welfare, as one interviewee put it:

‘Big issues have not been brought to my attention. We’re not out there killing patients.’

Although all interviewees believed that monitoring had improved study documentation, most were unsure whether monitoring had improved subjects’ rights and welfare: only one interviewee said that compliance monitoring had improved adherence to inclusion and exclusion criteria, and another interviewee said their quality-improvement programme had improved the informed consent process:
'A lot of people worry a lot about the form, and ... we've tried ... to ... educate them about how important ... the process [is] of how you talk to people when you approach them.'

However, only one monitor, of a quality-improvement programme, had actually measured the quality of studies before and after monitoring visits to substantiate claimed positive effects of monitoring, and found that study documentation had improved.

Some compliance monitors thought that monitoring had disturbed the relationship between the IRB and investigators, as one interviewee phrased it: 'Some researchers may feel that [the IRB is] ... policing them.' Negative effects of monitoring reported by many interviewees of both types of programmes were the investment of time by investigators and the use of institutional resources. For one institution, lack of resources was reason not to monitor at all.

3.4 Discussion

To summarize how ten US medical institutions monitored human subjects research: although monitoring programmes differed in many respects, they gravitated towards two general types – 'compliance monitoring' and 'quality-improvement monitoring'.

Compliance monitors were part of an IRB office, and visits were called 'audits'. The objective was to ensure compliance with regulations. Investigators could not decline participation. Monitors focused on documentation. Reports of visits were sent to the IRB, which could request mandatory corrective actions or take punitive measures.

Quality-improvement monitors were part of an independent office, and visits were called 'reviews'. The objective was to improve the research process. Investigators requested visits. Monitors focused more on actual research conduct. Confidential reports of visits provided voluntary recommendations, and were used to improve IRB review and institutional research policies and facilities.

Although several authors have reported on monitoring activities that were part of research projects, only VandenBosch and Miao have described regular monitoring programmes at US institutions (2011). Furthermore, monitoring activities described in
three Canadian studies appear to be similar to the activities described in this study, except McCusker et al., who reported that research subjects were routinely interviewed (Lavery et al. 2004, McCusker et al. 2001, Norton and Wilson 2008). Our finding that monitoring programmes vary between institutions is also in line with the study by VandenBosch and Miao (2011). Our distinction between a compliance philosophy and a quality-improvement philosophy is supported by other authors on human subject protection at medical institutions (De Jong 2011, Koski 2003, Noah 2004, Weijer et al. 1995). Moreover, we believe this distinction fits a much broader distinction between what sociologists call ‘civic’ and ‘industrial’ worlds (Boltanski and Thévenot 2006). Our finding that institutions use a compliance philosophy or a quality-improvement philosophy to monitor research can thus be explained by the fact that institutions have both a clear social responsibility and are part of the research industry.

Our study has two limitations. Since our findings reflect what interviewees reported about programmes, observing monitoring activities in practice might have revealed additional or incongruous information. Furthermore, although we used a mixed dataset, and our findings were supported by interviews with scholars not personally involved in the monitoring programmes in this study, the validity of our conclusions rests on a relatively small dataset.

Although it is possible that institutions in other parts of the US and institutions without federal funding use different types of monitoring, we have no reason to believe that our findings do not apply to many other US institutions. Furthermore, our findings might be relevant to institutions in other countries which require monitoring: European Union guidelines require that sponsors monitor the conduct of clinical trials, which implies that if an institution is the sponsors, e.g. investigator-initiated studies, the institution should monitor; (The Commission of the European Communities 2005) and regulations in Canada, Australia, and New Zealand require that RECs monitor (Douglass et al. 1998, McNeill et al. 1990, Weijer et al. 1995). Because the practice of medical ethical review is quite similar in the US, the United Kingdom, Canada, Australia, and New Zealand (Fitzgerald et al. 2006), we suggest that monitoring programmes could also be similar.

Our findings provide institutions with an overview of the possible ways of setting up a programme in order to meet the ethical and regulatory requirement to
monitor medical research. Furthermore, our findings highlight two important considerations for deciding how to set up a programme.

First, institutions have to consider to what extent they want to monitor. Our results lend some support to two reasons to monitor, discussed in the literature (Berry 1997, Bortolussi and Nicholson 2002, Christakis 1988, Douglass et al. 1998, Fitzgerald et al. 2006, Heath 1979, Hoffman 2001, Linden-Laufer 1997, McCusker et al. 2001, Norton and Wilson 2008, Smith et al. 1997, Weijer et al. 1995). The first reason is that monitoring protects the rights and welfare of subjects. Most interviewees in our study were not sure whether monitoring had improved subjects’ protection. However, monitors of both types of programmes thought that monitoring had improved study documentation. We think it is not unreasonable to expect that carefully executed studies have well-organized study documentation, and therefore that study documentation might function as a ‘surrogate outcome measure’ for human subjects protection. However, we also think that the relationship between study documentation and human subjects protection, and generally spoken, finding the most effective approach for protecting human subjects, warrants further investigation in order to prevent monitoring from becoming merely an additional layer of bureaucracy. The second reason to monitor is that it helps educate investigators about ethical research conduct. We found, indeed, that monitors considered education an important task. Again, although it is plausible that this contributes to human subjects protection, it was unknown whether this was the case.

Our results also support an important reason against extensive monitoring: the investment of time by investigators and use of institutional resources. We therefore suggest that in addition to considering effectiveness, institutions also take cost-effectiveness into account when deciding to what extent they will monitor. Our study shows one important way to maximize cost-effectiveness: select studies based on risk factors (table 3), so that only those studies are monitored for which subjects’ protection is needed most.

The second important consideration for setting up a monitoring programme which is highlighted by our findings, concerns the kind of programme. We have shown that institutions deploy a broad range of monitoring activities, which can be explained by the lack of guidance for institutions (VandenBosch and Maio 2011). In our opinion, a
lack of detailed official guidance is not necessarily problematic, since it allows institutions to set up programmes that suit their specific needs, organizational structure, and research culture. We have also shown that there is a more fundamental choice to be made between embracing a compliance philosophy or a quality-improvement philosophy. Our results indicate that both types of programmes have specific drawbacks and advantages in addition to the more general reasons to monitor or not, described above.

The strong focus on documentation issues appears to be both an advantage and a drawback of compliance programmes. According to our interviewees, an advantage of compliance programmes is that enforcing regulatory requirements concerning study documentation (Table 4) helps to protect the institution from federal (e.g. FDA) involvement and legal liability. However, interviewees also acknowledge that it is unclear whether focusing on documentation is a good way to protect human subjects. Since quality-improvement programmes use additional methods (e.g. touring the research facility, talking to research staff) to monitor, they have more opportunities to offer such protection. So, the focus on documentation is also a limitation of the compliance approach.

Another drawback of compliance programmes is related to the fact that they are carried out by the IRB office and can amount to punitive measures. According to our interviewees, this can damage the atmosphere of trust between investigators and IRB. We believe that this, in turn, could destroy the ‘informal monitoring system’, i.e. people voluntarily helping the IRB with research oversight by identifying problems related to subjects’ protection (Levine 1980). Because quality-improvement monitoring is not carried out by the IRB and findings are not directly communicated to the IRB (except in cases of serious research misconduct), it can help foster the trust fundamental to the existence of the academic community (Levine 1980). Contrary to our findings, we suggest that in order to remain trusted by investigators, quality-improvement monitors should not (further) investigate suspected research misconduct affecting human subjects (i.e. for-cause audits). Furthermore, although some people believe that increased oversight on research conduct will increase public trust, this is not necessarily true for oversight that is perceived as policing, which is, according to our study, the case with compliance monitoring (Holmberg 2004, Weijer et al. 1995).
A specific advantage of the quality improvement programmes we studied is that although the findings of visits remain confidential, they are also used, in a non-identifiable way, as feedback: to the IRB in order to improve the review process, e.g. feedback that the IRB needs to communicate the regulatory background of their demands; and to the institution in order to improve research policies and facilities, e.g. feedback concerning inadequate study support staff. We believe that this helps to foster the atmosphere of trust between investigators and the IRB, and helps the institution to raise its standards for human subjects protection.

All in all, we believe that our study indicates that monitoring according to a quality-improvement philosophy is better suited to protecting human subjects than monitoring according to a compliance philosophy. We therefore recommend that if an institution wants to set up a monitoring program, this should be done according to a quality-improvement philosophy.
References


Two prognostic indicators of the publication rate of clinical studies were available during ethical review

Jean Philippe de Jong, Gerben ter Riet, Dick L. Willems

Abstract

Objective
To identify prognostic indicators of the publication rate of clinical studies, available to research ethics committees during review.

Study Design and Setting
Retrospective survival study of a random sample of 100 studies, approved by a Dutch academic research ethics committee, with follow up information by questionnaire and bibliographic searches. Multivariate Cox regression analysis of the association between publication rate and seven factors available during review: six study characteristics and the number of letters sent by the committee during review representing the length of the review process.

Results
Two factors were associated with publication rate: studies with possible therapeutic benefit to participants were less likely to be published than non-therapeutic studies (adjusted hazard ratio 0.16, 95% confidence interval 0.03 to 0.54); with every letter sent, publication was less likely (adjusted hazard ratio 0.46 per letter, 0.17 to 0.98). Possibly, studies with more than minimal burdens to participants were more likely to

---

1 This chapter was published as: De Jong, J.P., Ter Riet, G. & Willems, D.L. (2010) Two prognostic indicators of the publication rate of clinical studies were available during ethical review, *Journal of Clinical Epidemiology*, 63, 12, 1342-1350.
be published than studies with minimal burdens (adjusted hazard ratio 3.90, 1.03 to 16.64).

Conclusion
We identified two prognostic indicators of publication rate. After suitable replication, research ethics committees might explore using prognostic indicators like these to target study protocols at high risk for non-publication. Discussing the risk of non-publication with investigators could help prevent non-publication.

4.1 Introduction
For individuals participating in clinical studies, helping others by contributing to medical knowledge is an important motivating factor, also with regard to accepting the possibly burdensome consequences (Sugarman et al. 1998). However, several studies have shown that 18% to 81% of studies approved by research ethics committees (RECs) are not published in scientific medical journals (Decullier et al. 2005, Dickersin et al. 1992, Easterbrook and Matthews 1992, Hall et al. 2007, Pich et al. 2003, Sune-Martin and Montoro-Ronsano 2003, Turer et al. 2007, Von Elm 2008). Assuming that unpublished studies have little impact on medical knowledge, this fails to do justice to the participants’ motives and is also considered scientific misconduct (Chalmers 1990). To protect the interests of participants, RECs review study protocols before initiation and have the responsibility to balance the gain in knowledge and the burden on and benefit for participants. This implies that RECs should not allow studies to go unpublished (Savulescu et al. 1996). Also, the Helsinki declaration, article 30, implies that RECs take into account obstacles to publication (18th WMA General Assembly 1964). Nevertheless, RECs do not routinely monitor whether study results are published, and consider the possibility of non-publication only when sponsor contracts could obstruct publication (Mann 2002, McCormack et al. 2005, Pearn 1995).

If it were possible for RECs to predict which studies are unlikely to be published, they could take this into account while reviewing study protocols and try to prevent non-publication. Therefore, we investigated the association between a number of factors (as “prognostic indicators”), readily available to RECs during review, and the
rate of publication. Although eight previous studies have investigated the association between study characteristics and publication in cohorts of REC-approved studies (Decullier et al. 2005, Dickersin et al. 1992, Easterbrook et al. 1991, Hall et al. 2007, Stern and Simes 1997, Sune-Martin and Montoro-Ronsano 2003, Turer et al. 2007, Von Elm 2008), these studies do not provide RECs with information that allows them to predict publication during review for three reasons: (i) all previous studies investigated studies that had been “completed” (studies in which the nature of results was known or that followed all participants as planned), or even smaller subsets, (Sune-Martin and Montoro-Ronsano 2003, Turer et al. 2007, Von Elm 2008) like “completed randomised drug trials” (Von Elm 2008). We think that RECs should be concerned about all unpublished studies, including studies that are not completed; (ii) one study only performed a univariate analysis which does not allow for assessing the likelihood of publication using multiple factors simultaneously (Sune-Martin and Montoro-Ronsano 2003); (iii) several studies focused on publication bias and, consequently, included factors that can only be known at study completion, such as the statistical significance of results (Decullier et al. 2005, Dickersin et al. 1992, Easterbrook et al. 1991, Hall et al. 2007, Stern and Simes 1997). Obviously, if RECs want to consider the likelihood of publication during review by using prognostic indicators, they can only use information available at that time.

Our study had two main goals: (i) to determine the percentage of unpublished studies approved by the REC of a major Dutch academic medical center, and (ii) to identify factors associated with the rate of publication. Finally, we describe the characteristics of several unpublished studies.

4.2 Methods

Our study population consisted of study protocols that had been approved by a single major academic Dutch REC between January 1, 1997 and September 1, 2006. Because, according to Dutch regulation, multicenter studies should be reviewed in full by a single Dutch REC, and reviewed solely in terms of feasibility by other RECs, we excluded protocols not fully reviewed by the REC we studied. All protocols described
biomedical research questions, involved human participants, and required ethical review under Dutch law. For pragmatic reasons, we randomly selected 10 protocols a year, making 100 protocols in total (Figure 1). Under Dutch law, our study did not require ethical review. We obtained permission from principal investigators (PIs) to investigate their protocols.

Our outcome measures were: time to first publication after REC approval; percentage of initiated, unpublished studies; and adjusted hazard ratios (AHR) for 7 factors based on information available during ethical review (Table 2).

In September 2006, we mailed a questionnaire to PIs on the status of their study: initiated/not initiated, ongoing/completed, participant inclusion, actual publications, expected publications, and reasons for non-publication. This allowed us to identify studies that, according to PIs, had neither resulted in, nor would result in publication. If (after at least two reminders by mail or telephone) a questionnaire had not been returned before February 1, 2007, the study was excluded from further analysis. We used information provided by PIs on actual and forthcoming articles to locate these in MEDLINE, EMBASE, PsychINFO, and the Cochrane Library up to February 1, 2007, which also defined the end of follow-up. We extracted the information on study characteristics (Table 1 and 3) from protocols and accompanying forms used for review, except for “year of approval” and “REC letters”, which we obtained from the REC archives.

We defined “publication” as an original article containing study results, indexed in MEDLINE, EMBASE, PsychINFO, or the Cochrane Library. We defined “initiated” studies as studies that had included at least one participant, according to information provided by PIs. A study is “completed” if the results are known to the investigators. “REC letters” reflects the number of letters the REC sent to PIs during review up until approval – an indicator for the length of the review process. “Sample size” and “study duration” (time between inclusion of first participant and last observation of last participant) refer to projected quantities in protocols. If circumstances are actively altered to investigate effects on parameters, a study is characterized as “interventional” (as opposed to “observational”). “Participant burden” are the burdens and risks in addition to regular clinical treatment. We defined
Figure 1. Identification of Study Protocols Reviewed by a Major Dutch Academic REC: 1997-2006

2726 Protocols submitted for review

2542 Approved

183 Not approved

2117 Full review by investigated REC (eligible study population)

425 Full review by other REC

100 Selected (10/year)

80 Responders to questionnaire

20 Non-responders to questionnaire

75 Available for predicting model

5 Missing data on study characteristics
“participant burden” as “minimal” when participation involved venipunctures, noninvasive procedures like survey, diary, magnetic resonance imaging (MRI), electroencephalogram (EEG), electrocardiogram (EKG), ultrasound, physical examination, or required investment of time. Phase 1/2/3 drug trials, withdrawal of prescribed medication, diet, radiation exposure, and invasive procedures (except for venipuncture) amount to “non-minimal participant burden”. This distinction is in line with criteria used by Dutch RECs. “Therapeutic benefit” refers to studies in which the health of participants could benefit directly from study procedures. In “study phase 1” safety is assessed, in “study phase 2” efficacy, in “study phase 3” effectiveness compared to current treatment, and in “study phase 4” long term effects.

We used STATA 10.1 software for all statistical analyses. To investigate time to first publication and the percentage of unpublished studies, we used Kaplan-Meier survival analysis. We calculated the percentage of unpublished studies for those studies that were initiated, because only initiated studies can place a burden on participants. For each study, we defined the date of REC approval as start of follow-up (t=0), the date of publication as an event, and the period between t=0 and end of follow-up (February 1, 2007) as time at risk. This means we combined observations with different starting points of follow-up.

To determine the association between factors and the rate of publication, we constructed a multivariate Cox regression model with the rate of publication as the dependent variable and 7 factors as independent variables (marked in Table 1 with an asterisk). We selected the factors prior to analysis based on literature and knowledge of the review process. Because the factors drug, study type and therapeutic benefit were highly correlated (Spearman’s rank correlation coefficients: study type and therapeutic benefit: 0.77; study type and drug: 0.71; therapeutic benefit and drug: 0.51) we excluded drug and study type, based on the consideration that therapeutic benefit is an important issue during review: non-therapeutic studies have to meet additional requirements. We tested the proportionality assumption of the Cox model with STATA’s stptest, using a P-value of \( \leq 0.05 \). Year of approval violated this assumption. In order to adjust the other associations for the influence of year of approval we stratified for it, using two strata (1997-2001 and 2002-2006). Furthermore, we performed a stepwise backwards selection procedure (P-value to enter <0.25 and <0.15 to stay) to identify the best predictors for
publication. We added robustness by bootstrapping the final model 1000 times. The final model did not violate the proportionality assumption. We used Cox regression analysis to calculate AHRs and their 95% confidence intervals (CIs) for each variable. A higher AHR signifies a higher rate of publication associated with that particular factor. For this analysis, we did not exclude studies that were not initiated because this is information unavailable to RECs during review.

To determine whether study characteristics were associated with the rate of non-response to the questionnaire (selection bias), we performed a multivariate logistic regression analysis with response (yes/no) as the dependent variable and 10 independent variables: the seven we used for our Cox analysis of publication, and also the factors drug, study type and year of approval. Adjusted odds ratios close to the neutral value of 1 signify no selection due the corresponding factor.

4.3 Results

The REC we studied fully reviewed and approved 2117 study protocols between January 1, 1997 and September 1, 2006 (Figure 1). We investigated 4.7% (n=100) of this study population. The response rate to our questionnaire was 80% (n=80). Five studies had missing values on variables, so we based our Cox regression model on publication on 75 studies (total follow-up time 4474 months; range 6 to 115, median 63). Our survival analysis on first publication was based on 65 initiated studies (total follow-up time of 2993 months (range 8 to 113, median 42). The regression analysis with response as the binary dependent variable (to check for selection bias) showed that PIs of studies with non-minimal participant burden were less likely (odds ratio: 0.21) to respond (Table 1).

Table 1 describes our study population (n=100) in terms of characteristics of study protocols. We could assess the status of 79% (n=79) of study protocols on February 1, 2007 (one questionnaire lacked data on study status). At that time 65 studies
Table 1: Characteristics of Study Protocols

<table>
<thead>
<tr>
<th>Study Characteristic</th>
<th>Total (n=100)</th>
<th>Responders (n=80)</th>
<th>Non-responders (n=20)</th>
<th>Adjusted Odds Ratio (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median (10th-90th percentile range)</td>
<td>Median (10th-90th percentile range)</td>
<td>Median (10th-90th percentile range)</td>
<td></td>
</tr>
<tr>
<td>REC letters* (number of)</td>
<td>2 (1-3)</td>
<td>2 (1-3)</td>
<td>2 (1-3)</td>
<td>1.17 (0.47-2.93)</td>
</tr>
<tr>
<td>Sample size*</td>
<td>80 (14-1000)</td>
<td>80 (14-800)</td>
<td>75 (11-1750)</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td>Study duration* (months)</td>
<td>18 (3-36)</td>
<td>24 (3-36)</td>
<td>12 (4-39)</td>
<td>1.03 (0.98-1.08)</td>
</tr>
<tr>
<td>Study type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observational</td>
<td>51 (51)</td>
<td>39 (49)</td>
<td>12 (60)</td>
<td>6.64 (0.54-82.30)</td>
</tr>
<tr>
<td>Interventional</td>
<td>49 (49)</td>
<td>41 (51)</td>
<td>8 (40)</td>
<td></td>
</tr>
<tr>
<td>Funding*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonprofit</td>
<td>65 (65)</td>
<td>52 (65)</td>
<td>13 (65)</td>
<td>0.94 (0.21-4.11)</td>
</tr>
<tr>
<td>Profit</td>
<td>34 (34)</td>
<td>27 (34)</td>
<td>7 (35)</td>
<td></td>
</tr>
<tr>
<td>Mono/multicenter*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monocenter</td>
<td>64 (64)</td>
<td>53 (66)</td>
<td>11 (55)</td>
<td>0.37 (0.09-1.52)</td>
</tr>
<tr>
<td>Multicenter</td>
<td>36 (36)</td>
<td>27 (34)</td>
<td>9 (45)</td>
<td></td>
</tr>
<tr>
<td>Therapeutic benefit*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No therapeutic benefit</td>
<td>52 (52)</td>
<td>39 (49)</td>
<td>13 (65)</td>
<td>3.02 (0.46-19.69)</td>
</tr>
<tr>
<td>Therapeutic benefit</td>
<td>46 (46)</td>
<td>39 (49)</td>
<td>7 (35)</td>
<td></td>
</tr>
<tr>
<td>Participant burden*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimal</td>
<td>39 (39)</td>
<td>33 (41)</td>
<td>6 (30)</td>
<td>0.21 (0.05-0.93)</td>
</tr>
<tr>
<td>Non-minimal</td>
<td>59 (59)</td>
<td>45 (56)</td>
<td>14 (70)</td>
<td></td>
</tr>
<tr>
<td>Drug</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>64 (64)</td>
<td>52 (65)</td>
<td>12 (60)</td>
<td>0.29 (0.03-2.82)</td>
</tr>
<tr>
<td>Yes</td>
<td>36 (36)</td>
<td>28 (35)</td>
<td>8 (40)</td>
<td></td>
</tr>
<tr>
<td>Study phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>64 (64)</td>
<td>50 (63)</td>
<td>14 (70)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2 (2)</td>
<td>2 (3)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>16 (16)</td>
<td>13 (16)</td>
<td>3 (15)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12 (12)</td>
<td>9 (11)</td>
<td>3 (15)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6 (6)</td>
<td>6 (8)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Topic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Therapy</td>
<td>47 (47)</td>
<td>37 (46)</td>
<td>10 (50)</td>
<td></td>
</tr>
<tr>
<td>Medical device</td>
<td>6 (6)</td>
<td>6 (8)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Diagnostics</td>
<td>26 (26)</td>
<td>23 (29)</td>
<td>3 (15)</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 4: PUBLICATION RATE OF CLINICAL STUDIES

<table>
<thead>
<tr>
<th>Category</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgical procedure</td>
<td>2 (2)</td>
<td>2 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Prevention</td>
<td>8 (8)</td>
<td>8 (10)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Basic research</td>
<td>23 (23)</td>
<td>16 (20)</td>
<td>7 (35)</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>60 (60)</td>
<td>49 (61)</td>
<td>11 (55)</td>
</tr>
<tr>
<td>Non-placebo controlled</td>
<td>19 (19)</td>
<td>13 (16)</td>
<td>6 (30)</td>
</tr>
<tr>
<td>Placebo controlled</td>
<td>21 (21)</td>
<td>18 (23)</td>
<td>3 (15)</td>
</tr>
</tbody>
</table>

Not all figures add up to 100 because some questionnaires were incomplete or archived protocols lacked information.

* Study characteristics marked with an asterisk were used in our model on publication.
† Multiple logistic regression model with response (yes/no) as the dependent variable and study characteristics as independent variables. Adjusted odds ratio>1 indicates a positive association with response.

Figure 2: Status of Study Protocols, February 1, 2007
Figure 3: Non-publication of Initiated Studies

had been initiated, of which 47 had been completed and 18 were still ongoing (Figure 2). Twenty-three studies resulted in one or more publications. Ten studies resulted in 1 publication; three in 2; three in 3; one in 4; two in 5; two in 6; one in 8; and one in 10; making a total of 69 publications. Two investigators reported only publications that did not satisfy our definition of “publication”. Forty-two initiated studies had not been published. In 33 of these, PIs still expected to publish. This left nine studies that had included participants, but that, according to the PIs, had neither resulted in, nor would result in publication. Fourteen studies had not been initiated (eight had been cancelled and six not yet initiated).
**Table 2:** Factors Associated with the Rate of Publication, multivariate analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Initial Model</th>
<th>Final Model †</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>REC letters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per additional letter</td>
<td>0.49 (0.22-1.10)</td>
<td>0.46 (0.17-0.98)</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 80 (ref.)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>&gt;80</td>
<td>0.55 (0.16-1.86)</td>
<td></td>
</tr>
<tr>
<td><strong>Study duration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per additional month</td>
<td>1.01 (0.97-1.05)</td>
<td></td>
</tr>
<tr>
<td><strong>Funding</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonprofit (ref.)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Profit</td>
<td>1.79 (0.63-5.07)</td>
<td></td>
</tr>
<tr>
<td><strong>Mono/multicenter</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monocenter (ref.)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Multicenter</td>
<td>0.97 (0.15-6.26)</td>
<td></td>
</tr>
<tr>
<td><strong>Therapeutic benefit</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (ref.)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.16 (0.05-0.52)</td>
<td>0.16 (0.03-0.54)</td>
</tr>
<tr>
<td><strong>Participant burden</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimal (ref.)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Non-minimal</td>
<td>2.91 (0.76-11.23)</td>
<td>3.90 (1.03-16.64)</td>
</tr>
</tbody>
</table>

† After backwards stepwise selection with P-Value <0.25 to enter and <0.15 to stay.
‡ Adjusted hazard ratio>1 indicates a positive association with publication.
ref. = reference category.

Our survival analysis shows that time to first publication of initiated studies (n=65) is at least 13 months and at most 60 months, at which point the percentage of unpublished studies is 46% (95%CI, 30-61) (Figure 3).

Table 2 shows the results of our multivariate Cox regression analysis with rate of publication as the dependent variable and 7 factors as independent variables (n=75). Backwards stepwise selection excluded the factors sample size, study duration, funding and mono-multicenter, leaving three factors in our final model: REC letters, therapeutic benefit and participant burden. With every additional letter sent during review by the REC to the PI, a study is 2.2 times less likely (adjusted hazard ratio 0.46; 95%CI 0.17-0.98) to be published. Studies in which participants could receive therapeutic benefits
Table 3: Characteristics of 9 Unpublished Studies†

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>REC Letters</th>
<th>Therapeutic Benefit</th>
<th>Participant Burden</th>
<th>Possible Benefits to Participants</th>
<th>Reason for Non-publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>2 Yes</td>
<td>Non-minimal (Burdens: physical examination, 3 abdomen/thorax CT scans, 30 visits to clinic, 29 venipunctures)</td>
<td>Non-minimal (Burdens: physical examination, 3 abdomen/thorax CT scans, 30 visits to clinic, 29 venipunctures Risks: 3 months of drug side effects: nausea, vomiting, gastralgia, hair loss, hepatotoxicity, seldom leucopenia)</td>
<td>Treatment of worm infection with variable morbidity</td>
<td>Insufficient inclusion of participants</td>
</tr>
<tr>
<td>2</td>
<td>59</td>
<td>2 Yes</td>
<td>Non-minimal (Burdens: more frequent hospital visits, &gt;4 venipunctures)</td>
<td>Non-minimal (Burdens: more frequent hospital visits, &gt;4 venipunctures Risks: drug side effects: nausea, headache, asthenia, diarrhea, abdominal pain, rash, sleeplessness, dizziness, vomiting, unknown drug side effects)</td>
<td>Improved treatment of severe chronic infection</td>
<td>Unknown</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>2 Yes</td>
<td>Non-minimal (Burdens: 6 surveys, 1 venipuncture)</td>
<td>Non-minimal (Burdens: 6 surveys, 1 venipuncture)</td>
<td>Symptom reduction in incurable malignancy</td>
<td>Toxicity in experimental group</td>
</tr>
<tr>
<td>4</td>
<td>1300</td>
<td>2 Yes</td>
<td>Non-minimal (Burdens: 18 hospital visits, 18 physical examinations, EKG, 18 venipunctures)</td>
<td>Non-minimal (Burdens: 18 hospital visits, 18 physical examinations, EKG, 18 venipunctures Risks: unknown drug side effects)</td>
<td>Improved treatment of cardiac risk factor</td>
<td>Negative results of interim analysis Sponsor stopped development</td>
</tr>
<tr>
<td>5</td>
<td>Unknown</td>
<td>3 Yes</td>
<td>Minimal</td>
<td>None</td>
<td>Measurement technique unreliable</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>3 No</td>
<td>Minimal</td>
<td>None</td>
<td>Results uninteresting</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>70</td>
<td>2 No</td>
<td>Minimal</td>
<td>None</td>
<td>Insufficient inclusion</td>
<td></td>
</tr>
</tbody>
</table>
### CHAPTER 4: PUBLICATION RATE OF CLINICAL STUDIES

#### Problems with research methodology

<table>
<thead>
<tr>
<th>Study</th>
<th>Burdens</th>
<th>Risks</th>
<th>Changes</th>
<th>Negative results of interim analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 8 3</td>
<td>Yes</td>
<td>Non-minimal (Risks: possibly an additional 1.5% risk of severe bleeding complication)</td>
<td>Reduction of organ failure in children</td>
<td></td>
</tr>
<tr>
<td>9 8 2</td>
<td>Yes</td>
<td>Non-minimal (Burdens: 7 hospital visits, 8 weeks withdrawal from pain relieving medication, physical examination, 9 surveys, 2 venipunctures, 1 IV injection, 2 thorax scans, 4 esophageal manometer procedures, 4 times 48-hour pH monitoring, 6 gastroscopies)</td>
<td>Reduction of chronic medication with mild side effects</td>
<td>Sponsor stopped development</td>
</tr>
</tbody>
</table>

† Initiated studies that will not result in publications according to principal investigators.

are 6.1 times less likely (adjusted hazard ratio 0.16; 95%CI 0.03-0.54) to be published than studies without therapeutic benefit. Studies with more than minimal burdens to participants were 3.9 times more likely to be published than studies with minimal burdens (adjusted hazard ratio 3.90, 95%CI 1.03-16.64).

1492 participants had been included in nine studies that, according to the PIs, had neither resulted in, nor would result in publication (Table 3). The burden on participants ranged widely between these studies. Some studies placed no burden at all on participants, others were time-consuming, involved a relatively mild burden, or drug side effects. Some studies involved more serious risks and burdens. On the other hand, several studies also involved possible therapeutic benefits for participants. Eight PIs provided reasons for non-publication, five of them possibly related to publication bias.
4.4 Comment

We found that an initiated study will probably remain unpublished if it is not published within five years after approval, at which point 46% remains unpublished. If we want to apply this figure to our total study population (n=2117) we have to allow for a percentage of studies that will not be initiated (in our sample up to 18%) and we have to bear in mind that our figure of 46% has a 30-61% confidence interval. Still, we can conclude that hundreds of studies approved by this single research ethics committee (REC) between 1997 and 2006 are likely to remain unpublished. Furthermore, 1492 participants had been included in nine studies, positively identified as unpublished. Some participants had been exposed to important health risks, whereas others possibly benefited. It is likely that a considerable part of these participants experienced a net health disadvantage. Combining these results, we tentatively conclude that many participants have been burdened by studies that will never be published.

Other studies have shown that 18% (Pich et al. 2003) to 81% (Dickersin et al.) of REC-approved studies are not published (Decullier et al. 2005, Dickersin et al. 1992, Easterbrook and Matthews 1992, Hall et al. 2007, Pich et al. 2003, Sune-Martin and Montoro-Ronsano 2003, Turer et al. 2007, Von Elm 2008). This broad range can be attributed to variation in REC location, duration of follow-up (ranging from 3 to 18 years), period of follow-up (ranging from before 1980 to 1999), and selection criteria used to identify study populations (ranging from complete cohorts of approved studies to only completed randomised drug trials). To our knowledge, only Decullier et al. and Easterbrook et al. provide data on the percentage of unpublished studies in a complete cohort (i.e. including all studies approved during a certain period) of approved studies (in France in 1994 and in the UK in 1984-1987, respectively). Decullier et al. found that eight years after approval, 67% of initiated studies had not been published, whereas Easterbrook et al. found that after 2.5 to 6.5 years, 64% had not been published. These figures are somewhat higher than our findings (46% after five years).

In our study, three factors were associated with publication: with every additional letter sent during review by the REC to the principal investigator (PI), a study was 2.2 times less likely to be published; studies in which participants could receive therapeutic benefits were 6.1 times less likely to be published than studies
without therapeutic benefit; studies with more than minimal burdens to participants were 3.9 times more likely to be published than studies with minimal burdens.

Five previous studies investigated factors associated with publication in a multivariate analysis of a complete cohort of approved protocols (Decullier et al. 2005, Dickersin et al. 1992, Easterbrook et al. 1991, Hall et al. 2007, Stern and Simes 1997). Although they restricted the analysis to the subclass of “completed” studies, they provide the best comparison for our results. They studied three factors identical to ours: sample size, funding (nonprofit/profit) and monocenter/multicenter. Although these factors were excluded from our final model due to backwards stepwise selection, our findings in the initial model are consistent with theirs, considering the 95% confidence intervals. Unfortunately, the previous studies did not investigate associations between publication and the number of letters sent during review, possible therapeutic benefit to participants, nor burdens on participants, so we cannot compare our findings on these three factors.

Although we investigated a small sample (4.7%) of our total study population, we think our findings may generalize well. The academic REC we investigated reviews protocols similar to the ones reviewed by other Dutch RECs, except for Phase 1 drug studies, which are reviewed predominantly by non-academic RECs (Centrale Commissie Mensgebonden Onderzoek (CCMO) 2007). Furthermore, our study population seems similar to previous studies with respect to sample size, study duration, observational/interventional, funding, monocenter/multicenter, topic, and control groups (Decullier et al. 2005, Dickersin et al. 1992, Easterbrook et al. 1991, Easterbrook and Matthews 1992, Hall et al. 2007, Stern and Simes 1997, Turer et al. 2007). Our study population contained somewhat fewer blinded study designs, and also fewer drug and Phase 1 studies (Decullier et al. 2005, Turer et al. 2007). We provide the most recent data (median year of approval: 2000).

Our study is limited by the small number of observations (75, with 20 events (publications) and relatively large number of associations (7) studied. Therefore, we refrained from exploring interactions between variables although such an exploration may prove fruitful in larger samples. We suggest interpreting the associations we found as tentative and in need of replication. Another issue is the limited amount of follow-up time (minimally 6 months) for some studies, whereas we show that the time from
approval to first publication ranges from 13 to 60 months. This means that some of the studies sampled did not have sufficient time to publish results. Although limited amount of follow-up time leads in a Kaplan-Meier survival analysis and Cox regression analysis to loss of precision, the loss to follow-up was not related to the outcome, because we ourselves limited the follow-up to February 1, 2007, which was the preplanned date of study termination. Therefore, the limited amount of follow-up time for some studies could not have affected the validity of our results (Rothman et al. 2008). A related issue is that we focus on information available at the time of ethics review. Therefore our model could not incorporate factors such as completion of the study or the significance of a study’s results, which are known to be strong factors affecting the rate of publication. Also, due to high correlation, we had to exclude the factors drug and study type. So the associations we found are only valid as predictors at the time of ethics review and should not be considered the only predictors of publication. A further note on the associations we found is that we had to stratify for year of approval to satisfy the “proportionality” assumption of the Cox model (the hazard of the event in any group should be a constant multiple of the hazard in any other) (Bradburn et al. 2003). This means that although the three factors in our final model have been adjusted for the year of approval, our data do not show whether the publication rate has changed over time. Furthermore, non-response was 20%. PIs of studies with non-minimal participant burden were less likely to respond. If the non-response of studies with non-minimal burdens was related to the fact that they had not been published, our finding that burdensome studies are more likely to be published would be exaggerated. Therefore, the association of burden with publication can be biased and should be interpreted with considerable reserve. Another limitation is that our data on burdens and benefits for participants in unpublished studies is based on expectations expressed in protocols, not on actual health outcomes. Finally, our definition of publication, as an article which can be found in MEDLINE, EMBASE, PsychINFO, or the Cochrane Library could be criticised. We consider publication in a peer-reviewed scientific journal, indexed in these databases as the “gold standard” of publishing study results. However, publication in other media (e.g. theses, local journals, guidelines, reports, presentations) also has an impact on the scientific community, which would make our definition too strict. On the other hand, it can be
argued that even publication in a peer-reviewed indexed scientific journal is not the hallmark of successful publication, and that impact factors, citations, the quality of the study, or having actual clinical consequences should be taken into account, which would make our definition too lenient.

Because we have restricted our model to seven factors known at time of ethics review, our results should be interpreted as “prognostic indicators” and not directly in terms of causality. We can only speculate on the causal mechanisms behind the association of the rate of publication with (i) the number of REC letters, (ii) therapeutic benefit, and (iii) participant burden. (i) Ethical review can be difficult due to procedural factors (e.g., lack of clarity in the protocol, imprecision in the forms, or not fully answering the REC’s questions), to more substantial factors (e.g., lack of scientific value, methodological inadequacies), and to the complexity of the study protocol. These factors lead to a longer review process with an increase in correspondence between REC and investigators, which is reflected by the factor REC letters. We suggest that the same factors that cause an increase in correspondence may also cause difficulties in completing and publishing the study. (ii) As for the association with therapeutic benefit a causal mechanism might lie in publication bias. It seems plausible that therapeutic studies which aim to benefit participants directly suffer more from biased (i.e. diminished) reporting, because in this type of studies ‘negative’ and ‘non-significant’ results can be more clearly distinguished from ‘positive’ than is the case with more fundamental studies. A second candidate for explaining this association lies in the ethical review process. It seems plausible that studies which lack the possibility of benefiting participants are reviewed more strictly, which could increase the studies’ quality and consequently, the rate of publication. (iii) An encouraging speculation on the finding that burdensome studies are associated with a higher rate of publication would be that in these studies investigators are more motivated to publish because they feel responsible for burdening participants. However, as explained before, this association should be interpreted with reserve.

To help prevent non-publication of study results, RECs could consider taking the likelihood of publication into account while reviewing studies. Since we have found factors associated with the rate of publication, and have found that a considerable part of studies will probably remain unpublished, we conclude that RECs may have early
indicators of non-publication. Their use for individualized predictions has to be investigated in a larger study. Still, even if it would be possible to predict non-publication with great confidence, the question remains as to what a REC should do with such information. Withholding approval based on a prediction is problematic for two reasons: (i) Using study characteristics to approve or disapprove a study could introduce incentives to alter these characteristics. Since we have only established associations and not causal relationships, this would not automatically lead to more publications. (ii) RECs should provide good reasons for disapproving a study. The mere fact that a study has certain (in themselves unobjectionable) characteristics cannot justify disapproval. We think the only just way for a REC to use factors associated with the likelihood of publication is to use them as indicators to further scrutinize a protocol and to discuss their concerns about non-publication with investigators. In any case, we think that when weighing up the burden on participants and the gain in knowledge, a REC should bear in mind that on average a study has only about a fifty-fifty chance of contributing to medical knowledge.

A second way RECs might deal with non-publication is to monitor ongoing studies and sanction non-publication. RECs could demand an adequate motivation for refraining from publishing study results, could denounce non-publication, or even restrict future research activities. Although many RECs are legally required to monitor studies through the investigators’ annual and final reports, this generally does not include monitoring publication (Mann 2002, McCormack et al. 2005, Pearn 1995). We believe that the merits of monitoring studies to increase publication deserve further exploration.

Although we acknowledge that it would not be fair to condemn the nine unpublished studies we found, based on the limited information provided by PIs (table3), we think more studies could and should be published – definitely more than half – and think that having negative results, losing scientific interest, and lack of time are not in themselves valid reasons to refrain from writing down the results and submitting them to a journal (Dickersin et al. 1987, Dickersin and Min 1993, Weber et al. 1998). On the other hand, we acknowledge that some studies can fail to yield meaningful results because of unforeseen problems with inclusion of participants, research methodology and lack of funding. Still, even “failed” studies can be of scientific
interest to other researchers. Since we also appreciate that journals do not want to publish studies they consider inadequate, we suggest that a part of the solution here could lie in trial registries that can provide valuable information on “failed” studies (Dickersin and Rennie 2003). Furthermore, to account for the inevitability that some studies will not be published, we believe RECs could easily adopt a third way to do justice to participants: just as physicians warn their patients about possible failure of treatment, have investigators warn participants during informed consent that their study might not be published and that, consequently, participation might not contribute much to the advance of medical science (Savulescu et al. 1996).
References

18th WMA General Assembly (1964) World Medical Association Declaration of Helsinki; Ethical Principles for Medical Research Involving Human Subjects.


TUMOUR TISSUE: WHO IS IN CONTROL?1

Guidelines on tissue banking for clinical and research purposes

Abstract

Recent developments in genomics have resulted in the increased availability of gene profiles for early diagnostics and prognostics in breast cancer. We expect that genetic analysis of a patient’s (tumour) tissue will, in time, become a standard part of cancer treatment. A request from a Dutch woman to have her tumour tissue tested years after treatment confronted the Netherlands Cancer Institute (NKI) and its staff with legal, ethical and practical questions regarding patients’ rights in relation to residual tissue storage and its use for clinical purposes. Was her tissue still available? If so, could she (and her relatives) demand that the test be carried out? Or, could she demand that the tissue be transferred to another hospital? As it became apparent that appropriate guidance was lacking in this area, the NKI arranged for a group of professionals with legal and ethical expertise to develop a guideline within the framework of a Technology Assessment project. Subsequently, the relevant stakeholders, including oncologists, pathologists, medical researchers and patient representatives, were invited to reflect on the guideline, including its practical implications. Consensus was reached on the

guideline, including its main practical implications and the preservation of a sufficient amount of a patient’s residual tissue: exclusively for future use in diagnostics and prognostics. Finally, the staff of the pathology department was asked to report on the feasibility of the guideline given its current tissue banking procedures.

5.1 Introduction

A promising technology in the treatment of breast cancer is the introduction of genomic profiling for prognostics and (early) diagnostics. This development puts the availability of tissue for clinical care in the spotlight. We expect that both hospitals and medical professionals will increasingly be confronted with legal and ethical questions concerning the rights of patients regarding their tissue. Can a patient expect that enough suitable tissue is stored for his future treatment or that of his relatives? Should a patient, after initial treatment, be informed about newly introduced diagnostic or prognostic tests, and if so, to what extent? Can a patient demand that a test be performed and/or that his tissue be transferred to another hospital? Likewise, what are the rights of a patient’s relatives with regard to that patient’s tissue?

While a number of issues concerning the storage and use of tissue for research purposes, such as ownership and informed consent, have already been extensively discussed in the literature, little focus has been placed on patients’ rights regarding tissue banking for clinical purposes (Hansson et al. 2006, McHale et al. 2007). Against that background, the Netherlands Cancer Institute (NKI) initiated a research project covering these questions. The project resulted in a guideline on tissue storage and use for clinical purposes, specifically focusing on patients’ rights. This paper concentrates on the guideline, its underlying principles, its main provisions and the feasibility of its application in clinical practice. In the following, first the background to the guideline is briefly described. Then we address the fundamental legal and ethical principles underlying the guideline and its main provisions, which should, in our opinion, form the basis for more detailed institutional regulations. Finally, we make some observations about the feasibility of applying the guideline in hospital practice, making reference to the first experiences of the NKI’s pathology department.
5.2 Background to the guideline

5.2.1 Promising developments in genomics
Gene-expression profiling is an important development that is likely to predict the diagnosis and prognosis of malignant disease more accurately than existing clinicopathological parameters (Bertucci et al. 2001). Although gene-expression profiling is not yet routine procedure, several tests are currently under investigation. In the United States for example, the prognostic and predictive accuracy of a genomic profile for breast cancer patients (called the 21-gene recurrence score), based on paraffin embedded tumour tissue is currently being studied in a randomized trial (Paik et al. 2004). In Europe, a 70-gene signature (MammaPrint), which uses microarrays on fresh frozen tumour tissue, is being tested in a multicenter randomized trial (MINDACT) for its prognostic and predictive accuracy (Bueno-de-Mesquita et al. 2007, van ’t Veer et al. 2002, van de Vijver et al. 2002). Although genomic profiling will in the first place help make cancer treatment more effective, in the future genetic profiling will probably be used for many other diseases and for other goals than prognostics, such as for establishing the presence or absence of a disease or its responsiveness to therapy. For the successful performance of these tests tissue should be available, both in a proper form and in sufficient quantity.

5.2.2 Developing a guideline
In the course of the feasibility and technology assessment study of the 70-gene-signature in the Netherlands, it became apparent that implementation of these new diagnostic and prognostic technologies generates new legal and ethical questions concerning the storage and use of a patient’s tissue for clinical purposes (Douma et al. 2007). We refer here to the questions addressed in the introduction. Further encouraged by an actual request by a Dutch woman previously treated for breast cancer to have her tumour tissue tested with the 70-gene test, the hospital appointed a group of lawyers and ethicists to study the new questions, together with the professionals concerned, i.e. physicians of the departments of oncology and pathology and researchers in the field of genomic profiling. Following the exploratory phase, a draft guideline was written and discussed with professionals and patient representatives.
during two subsequent meetings. Finally, the hospital’s department of pathology was asked to explore the feasibility of applying the guideline in daily clinical practice.

5.3 The guideline's underlying principles

The literature addresses tissue banking for clinical purposes much less than tissue banking for research purposes (Hansson et al. 2006). Since tissue banking for clinical purposes must also comply with international legislation and guidelines, we had to look at related legal and ethical documents. On an international level, our research included the Council of Europe’s Convention on Human Rights and Biomedicine and its additional protocols on Biomedical Research and Genetic Testing for Health Purposes, the Recommendation of the Council of Europe on research on biological materials of human origin and UNESCO’s International Declaration on Human Genetic Data; and on a national level we looked at regulations on patients’ rights, privacy protection and the use of tissue for research purposes (Council of Europe 1997, Council of Europe 2005, Council of Europe 2008, Council of Europe, Committee of Ministers 2006, Federation of Biomedical Scientific Societies 2001, UNESCO 2003, Wet bescherming persoonsgegevens 2000, Wet op de geneeskundige behandelingsovereenkomst 1994).

From these documents, we distinguish four general principles, which should be taken into account. First, care providers have a moral and legal obligation to protect the clinical interests of their patients (World Medical Association 1948). In light of emerging technologies, we think that good clinical care should include securing the availability of sufficient tissue for future clinical care and access for patients (and ultimately their relatives) to generally accepted diagnostic or prognostic tests on that tissue.

Second, irrespective of whether they can be considered “owners” of their removed tissue in their own jurisdiction, patients have personal rights regarding their removed bodily material (Charo 2006). We primarily aim to take into account the right of patients to consent or object to the storage and use of their tissue for other purposes than that for which it was removed, such as research (Council of Europe 1997). This implies that patients should be informed about the consequences of storage-and-use
policies of the hospital for themselves and their relatives. As patients have the final say over their residual tissue, a patient should always be able to request its destruction, unless he has agreed with its storage and use for research. As care providers are likely to differ in their policies on the availability of tests and the conditions under which they are accessible to patients, patients should also be entitled to request tissue transfer in order to have their tissue tested elsewhere.

A third principle concerns the position of the patient’s relatives. Here, the underlying question is whether a physician’s duty to provide good clinical care involves the protection of the relatives’ medical interests too. From the perspective of a physician’s duty to provide good clinical care, the mere fact that stored tissue can also serve health interests of genetically related relatives justifies protection of their position (Haites et al. 2001). Receiving genetic information of clinical relevance is an example of such an interest. Article 17 of the United Nations International Covenant on Civil and Political Rights (United Nations 1966) and Article 8 of the European Convention on Human Rights (Council of Europe 1950) point in this direction too, while they connect the constitutional right to private life with the recognition of the family as the basic unit. The Council of Europe’s Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research in Articles 13-15 specifically refers to situations in which tests could benefit relatives while the patient himself cannot provide consent (Council of Europe 2005). Finally, we refer to one of the conclusions of the Icelandic Supreme Court in its internationally discussed decision of November, 27 2003 (i.e. that not only patients themselves, but also their children or other relatives in the first line deserve privacy protection whenever their genetic data are being collected, stored and used) (Gevers 2004, Guðmundsdóttir v Iceland 2004). On the other hand, it is generally acknowledged that physicians have less extensive obligations to patients’ relatives than to patients themselves, as they are primarily responsible for the care of the persons who were seeking their assistance (Council of Europe 2008). This implies that as long as patients are capable of giving authorization, they should decide whether their tissue is tested or transferred in the interest of relatives.

The final principle that can be derived from international documents concerns the situation in which a patient’s interests conflict with the general interest of medical science. We refer to Article 2 of the Biomedicine Convention stating that “the interests
and welfare of the human being shall prevail over the sole interest of society or science” (Council of Europe 1997). A similar article is incorporated in UNESCO’s Declaration on the Human Genome: “No research or research application concerning the human genome (…) should prevail over respect for the human rights, fundamental freedoms and human dignity of individuals or, where applicable, of groups of people” (UNESCO 1997). Therefore, in situations in which tissue has been stored for the purpose of medical care as well as scientific research and is insufficient to serve both purposes, the medical interest of the patient overrides the general interest of doing scientific research. Only if a patient has been asked, has agreed and can oversee the consequences of his decision to donate his or her material specifically for research, does it seem reasonable that it may be used exclusively for that purpose (Winickoff and Winickoff 2003).

5.4 Main provisions of the guideline

5.4.1 Duties of hospitals and professionals

Primarily, excised tissue is sent to the pathologist for diagnostic purposes. In a number of cases it is difficult to obtain sufficient tissue and discussions on patient rights or scientific storage regulations should always take this into account. Once there is residual tissue, it should be acknowledged that patients have legitimate health interests in the availability of proper tissue.

The hospital’s primary responsibility in our opinion is to ensure that, as far as is reasonably possible, enough of a patient’s tissue is available for present or future clinical use. This implies that the practitioners responsible (i.e. surgeons and pathologists) should therefore ensure that sufficient tissue is stored and preserved in such a way – fresh-frozen or otherwise – that it is suitable for testing, even many years after initial diagnosis or treatment. In addition, the tissue should be stored as long as is necessary to serve the patient’s clinical interest, unless during that period the patient explicitly requests the destruction, donation or transfer of his tissue to another hospital. Because Dutch law does not regulate the storage period of human tissue, its length depends on professional and local guidelines. However, this may be different in other countries.
Second, after expiration of the required storage period, a hospital may destroy the tissue, but only if at that moment the medical interest of the patient, his relatives or the general interest of medical science no longer requires its retention.

Third, where the legal relationship between a patient and a hospital has come to an end and the patient or his relatives request that his tissue be transferred for diagnostic purposes, the hospital should cooperate with the transfer of the tissue to a different hospital. The right to have one’s tissue transferred to a different hospital should also apply when the attending practitioner refuses to do the requested test.

A final, but crucial, responsibility of the hospital is to develop local guidelines that cover all the relevant administrative aspects concerning tissue banking for clinical purposes, not in the least to be able to provide clear information to patients about their rights.

Apart from their previously cited responsibilities, physicians, who are primarily responsible for patient care, should consider it their duty to keep the patient informed about new clinically relevant tests that can be done on stored tissue as soon as these tests, according to local or national medical standards, can be considered an element of evidence based, good clinical care. We feel that it is a responsibility of professional organizations together with patient representatives to develop more detailed standards on what the responsibilities of physicians should entail in this respect. Making an exception only seems justified when it would be reasonably impossible for the physician to provide such information. This appears to be the case when the physician-patient relationship has ended or a patient is no longer traceable, or if he has invoked his right not to know. Once a test is generally accepted as standard practice and relevant, it should be offered to patients, or at least be carried out when patients demand it. If a situation occurs in which a patient and a physician disagree about the usefulness of a test, and a different physician is prepared to perform it, the tissue should be transferred at the patient’s request.

5.4.2 Rights of patients and their relatives
To ensure that patients are aware that their tissue is being stored for a long time and that they have an important say about what happens to it, they should receive adequate information on the storage period and use of their tissue, their personal rights and those
of their relatives. Without such information, the legal recognition of a patient's position with respect to his tissue remains theoretical. The rights of patients can further be incorporated in local rules for tissue banking (see the section above), and should include a right to the destruction or donation of the tissue and a right to have the tissue transferred to a different hospital. The latter right can be particularly instrumental in a situation where patients (or their relatives) are treated by a physician working in a different hospital, or when the attending practitioner refuses to perform the requested test.

With regard to the position of relatives, their interests should be protected to at least some extent. Subject to the condition that the patient concerned consents to it, relatives should also be able to request tissue transfer and testing related to their legitimate health interests. When a patient has died or is incapable of giving consent but the tissue is still available, relatives, in our view, have the right to request continuation of storage, transfer of the tissue to a different medical center, and testing of the tissue.

5.5 Some observations from practice

After all parties involved committed themselves to the guideline, the NKI's pathology department assessed the feasibility of incorporating the requirements of the guideline in their present tissue banking procedures.

The department reported that it is already NKI policy to store tissue for a practically limitless period. In situations of scarcity of the available tissue, it is not permitted to use it for research purposes, unless the patient consents to the donation of the tissue for that purpose. As to the actual application of the guideline's provisions, the department notes that it also has responsibilities towards researchers and has to support and facilitate tissue storage and use for clinical and retrospective research. From that perspective, it would be helpful to appoint a “tissue bank manager”, responsible for matters such as the further automatization of the record keeping of specimens and the assessment and handling of tissue. Furthermore, the actual selection, division and
preparation of the tissue should be conducted by well-trained pathologists and/or (senior) technicians.

Finally, provided it is reasonably possible, it would be preferable to store a “control piece” in a separate tissue bank, reserved exclusively for patient usage. These measures can have financial implications that should be taken into account.

A number of other issues remain to be addressed. Priority is currently given to the storage of paraffin embedded tissue, whereas fresh-frozen tissue could be more informative and promising for the future. However, this should not be an individual pathologist’s or hospital’s choice, but is an issue that should be discussed within the framework of professional and institutional obligations and of the principles set out above.

A different question is how long the obligation to store tissue should extend. In our view, this is primarily a matter of professional medical judgment as it depends on both the type of tissue and the significance and the expression modes of related prognostic or genomic factors. Genetic counselors are likely to advise a lengthy period, as relatives may benefit from a comprehensive “tissue-history” in future situations. On the other hand, pathology departments may consider limitless storage obligations neither reasonable nor feasible. Apart from this, privacy considerations could play a role, for instance with regard to minors (Gurwitz et al. 2009).

Although this guideline is primarily developed to inform storage policy on tumour tissue, we expect that it is also relevant to the storage of other types of tissue. We are aware that the presented elements require further reflection and debate. It is obvious that tissue storage for clinical purposes urgently needs further attention from medical, ethical, legal and practical perspectives. Hopefully, the guidance we propose will contribute to the discussion of this important issue.
References


Council of Europe Committee of Ministers (2006) Recommendation of the Committee of Ministers to member states on research on biological materials of human origin.


Abstract

The principle of justice is a crucial aspect in the ethical evaluation of medical research with human subjects. The principle requires that the burdens and benefits of participating in medical research should be distributed fairly among the patient population, but also that the benefits related to the results of medical research should be distributed fairly; the principle is also about whether relevant knowledge is available to aid the treatment of illnesses in different patient groups. In this article I explore the idea that the choice for a specific research methodology can affect how the benefits of research results are distributed within the patient population. Specifically, I suggest that the current ‘gold standard’ methodology for studying interventions, the randomized controlled trial (RCT), can, in particular cases, lead to a distribution of research results that is unfair to disadvantaged patient groups.

I argue that three characteristics of RCTs, all related to standardization, can potentially lead to an unfair distribution of research results: (1) RCTs tend to standardize study groups, and results in standardized study groups can be difficult to extrapolate to disadvantaged groups outside a study; (2) RCTs standardize the delivery of health care within studies, and disadvantaged patient groups outside a study may have more difficulties adapting to these standardized interventions; and (3) RCTs

1 Submitted for publication
require that interventions are standardized, which can mean that less relevant knowledge will be generated on more complex interventions such as lifestyle interventions, which are needed especially by disadvantaged patients.

I conclude that, although standardization in RCTs can help to generate valid knowledge, in particular cases it can also mean that less relevant medical knowledge is available for disadvantaged patient groups. This is an injustice that could be partly remedied by the efforts of good clinicians and by putting less emphasis on standardization in (particular) RCTs. In addition, I suggest a more radical solution to do justice to the multitude of differences between patients: stop considering RCTs to be the gold standard approach for the study of healthcare interventions, and turn in relevant cases to methodologies in which standardization is a less central concept, such as observational studies and especially qualitative methodologies. These are considerations that could be taken into account in ethical oversight.

6.1 Introduction

Besides the principles of respect for persons and beneficence, the principle of justice is a crucial aspect in the ethical evaluation of medical research with human subjects (Emanuel and Wendler 2000, Kahn et al. 1998). In this article, I argue that the choice for a specific research methodology can affect how the benefits of research results are distributed within the patient population. Specifically, I suggest that the current ‘gold standard’ methodology for the study of health care interventions, the randomized controlled trial (RCT), can in some cases lead to an unfair distribution of relevant knowledge on the treatment of illnesses in different patient groups, affecting disadvantaged patient groups in particular.

Important drivers for the attention to justice in medical research have been scandals involving vulnerable research subjects, such as the Nazi experiments on Jews and prisoners, and the Tuskegee syphilis study on poor African-American men. In the Belmont Report, the question of justice in research is addressed as follows: “Who ought to receive the benefits of research and bear its burdens? This is a question of justice in the sense of “fairness of distribution” (U.S. National Commission for the Protection of
Human Subjects of Biomedical and Behavioral Research 1978). In this article I define an ‘injustice’ or ‘unfairness’ (terms I use interchangeably) as an uneven distribution of relevant knowledge among people with medical needs; in such a way that it especially affects people who are already disadvantaged, for example in terms of the level of education, work and living conditions, income, social position/autonomy, psychological vulnerability, or as a rough measure that encompasses many of these disadvantages: socioeconomic status.

Regulations in many countries currently offer special protection to vulnerable groups such as prisoners, pregnant women and children in order to distribute the burdens of participation in medical research fairly among different patient groups (Emanuel and Grady 2007, McCarthy 1997). In the 1980s, access to research also came to be seen as a question of justice and some of the protections provided by regulations were increasingly perceived as overprotective: AIDS activists and other groups with high unmet medical need claimed that they were being treated unfairly because they were denied access to studies and could not benefit from participation in medical research (Emanuel and Grady 2007, Mastroianni and Kahn 2001).

These are examples of concerns about a fair distribution of the burdens and benefits related to participation in medical research. A parallel development has been a concern for a fair distribution of the benefits related to the results of medical research; so, whether research will lead to relevant knowledge to aid the treatment of illnesses in different patient groups (Allmark 2004). To correct for injustices in the distribution of relevant medical knowledge, research on diseases prevalent in specific groups has been stimulated in recent years through grants and regulations. Examples of this are research on rare diseases stimulated by ‘orphan drug’ policies, on diseases in developing countries under the header of ‘neglected diseases’, and on diseases more prevalent in specific groups, for example on diabetes in people with a lower socioeconomic status, and on hypertension in people of African descent. A recent overview of these developments is provided by the WHO Priority Medicines report 2 (Kaplan et al. 2013).

So, to date the concerns for justice in medical research have focused on who should receive the burdens and benefits of participation in research and what diseases should be the topic of research. However, in this article I will explore the idea that the choice of the research methodology is also relevant for justice in research. More
specifically, I will argue that the current ‘gold standard’ for studying health care interventions, the randomized controlled trial (RCT), can, in particular cases, lead to an unfair distribution of relevant medical knowledge. Currently, randomized controlled trials (RCTs) are widely considered to be the ‘gold standard’ of methods for studying health care interventions (Sackett et al. 1996, Timmermans and Berg 2003, p. 27): the bulk of medical research is conducted in the form of RCTs (Getz 2010); RCTs are considered to deliver the highest level of primary evidence for improving health care (Centre for Evidence-based Medicine 2013); and they are of key importance for regulating the market access of medicines (Rågo and Santoso 2008).

In the next three sections I will analyze three characteristics of the methodology of RCTs and argue that they can, in particular cases, lead to an unfair distribution of the benefits related to the results of research. In section 2, standard patients, I discuss how control and intervention groups are selected in RCTs. In section 3, standardized care, I discuss how RCTs affect the delivery of care by using protocols. In section 4, standardized interventions, I discuss what kinds of interventions can be studied with RCTs. I illustrate my argument by using examples of research on hypertension management. In section 5, I conclude by suggesting that medical research could lead to a fairer distribution of relevant knowledge if RCTs would not be considered the gold standard and if more studies would use observational, and especially qualitative, methodologies.

6.2 Standard patients

RCTs have a long history of ‘standardizing’ patients, which can lead to an unfair distribution of the benefits related to the results of medical research. I define ‘standardization’ as the process of making something conform to a fixed norm. One of the key aspects of the methodology of RCTs is dividing patients into two groups: an intervention group that is treated with the (new) health care intervention, and a control group that is treated with a control intervention (a placebo or regular care). Besides a difference in the interventions, the study groups need to be as much alike as possible because differences between groups can introduce bias and skew the outcome of a
study. Also to decrease ‘noise’ (random error) and increase the precision of measurements (Hill 1951), RCTs therefore have a tendency to focus on standard patients. In practice, this means that RCTs have often only studied the ‘40 year old white male’ (Epstein 2007), and excluded patients with ‘special’ characteristics (women, children, elderly and ethnic minorities). This practice of excluding these ‘special’ groups also fitted well with the ethical concerns about protecting vulnerable groups from research.

However, to be able to use the results of a study on ‘standard’ patients to guide decision making in regular clinical practice, health care providers needed to extrapolate from middle-aged white men to individual patients, which was not without risk. For example, it could turn out that women or children experienced severe side-effects much more often than the ‘standard’ patient. So, by conducting studies on standard patients, other groups, such as women, children, elderly and ethnic groups, benefited less from the results of research. For instance, although there are many adequately studied treatment options available to treat hypertension in men, there is still a lack of proper knowledge on antihypertensive medications for use with pregnancy-induced hypertension.

To remedy the injustice caused by a focus on standard patients, regulations and policy currently require that a study population is not restricted to middle-aged white men but represents the general patient population. In particular, the inclusion of women, children, elderly, and ethnic groups is encouraged (Epstein 2007, European Commission 2004). For example, for the evaluation of new anti-hypertensive medicines it is now recommended that “patients from relevant demographic subsets should be studied, including both men and women, racial/ethnic groups pertinent to the region, and both young and older patients” (ICH Steering Committee 2000). However, averaging the results of a study with a representative study population might make these results in theory more applicable to the ‘average’ patient, in clinical practice this average patient does not exist: it still requires an extrapolation from the study population to individual patients, with its accompanying risks. This is a problem that especially affects smaller subgroups (e.g. infants, ethnic minorities, pregnant women etc.) because data on them will be ‘swamped’ by those of larger patient groups. For example, it was only discovered after many years of use in clinical practice that beta-
blockers, one of the most frequently prescribed medicines for lowering blood pressure, are not effective in people of African descent (Brewster et al. 2004).

Although extrapolating from an average study population can in principle be problematic for every non-average patient group, I believe that it is especially problematic for disadvantaged patient groups. It is difficult to see how generally well-off groups such as the very rich, or people with a particularly high level of education would respond differently to an intervention (for example by suffering from more severe side-effects), compared to the average patient. This is supported by the fact that these groups do rarely receive special attention in clinical practice. On the other hand, it is a known fact that disadvantaged groups such as children, women, elderly, and ethnic groups can react differently to an intervention.

The acknowledgement of the problems with extrapolation has led to additional regulations and policies that require a separate analysis of subgroups to get insight in any relevant differences (Allmark 2004, Epstein 2007). This has been implemented by designating particular subgroups, defined by sex, age and ethnicity, as the relevant groups to be analyzed. This has led to valuable knowledge on how certain medicines affect particular disadvantaged groups, and in a more equal distribution of the benefits of medical research (Epstein 2007).

However, which patient groups are considered relevant has been standardized: sex, age and ethnicity are currently the standard subgroups. So, again, research subjects have been standardized, only now not according to a single standard but by using multiple standards, a phenomenon we could call ‘niche-standardization’ (Epstein 2007). Although I admit that ‘niche-standardization’ has helped to generate relevant knowledge for some of the disadvantaged groups, I argue that it still might not be doing justice to all disadvantaged patient groups. The current standard subgroups are not based on scientific evidence, but have been shaped by an alignment of classical biological distinctions with long-standing political categories (Epstein 2007) and are used irrespective of the disease or medicine under study. I think that subgroupings should be based on a thorough scientific analysis, taking into account factors such as genetic makeup, physiology and comorbidity, and that it should be decided on a case-by-case basis which ones are relevant. Furthermore, subgroupings based on psychological and sociological distinctions could turn out to be just as relevant for
particular medicines and diseases as biological subgroupings. For example, in the study of anti-hypertensive medication it could make sense to study subgroupings according to socio-economic class, since socio-economic class is strongly correlated to lifestyle factors known to affect hypertension, such as diet, body weight, activity and stress. Yet currently, relevant guidelines make no mention of those factors and focus on the standard biological subgroupings defined by sex, age and ethnicity (European Medicines Agency 2010).

I believe that it is possible to define other, non-standard subgroups and include these in RCTs in order to generate knowledge that is directly relevant to disadvantaged groups and thereby do better justice to these groups. However, setting up criteria for defining these groups, and generating hypotheses about whether they might be relevant in particular cases, would require using other types of methods besides RCTs, for instance observational methods. Furthermore, studying many kinds of subgroups would go against the strive in RCTs to reduce variation that can skew the outcome of a study or decrease the precision of measurements. Moreover, including additional subgroups could mean that a large number of research subjects is required to attain adequate statistical power, with the associated burdens and costs.

In sum, I have argued that RCTs have a long history of standardizing patients, and efforts to do better justice to subgroups of patients tend to (niche-)standardize patients. Because results in standardized study groups can be difficult to extrapolate to disadvantaged groups, this can lead in particular cases to an unfair distribution of relevant medical knowledge.

6.3 Standardized care

In this section I argue that the way interventions are used within the context RCTs – as standardized care – can lead to an unfair distribution of the benefits of medical research. RCTs assign patients randomly to either an intervention or a control group in order to ensure that physicians cannot distort an even distribution of patients between these groups. This process of randomization is part of broader efforts in RCTs to prevent variability in the delivery of health care from distorting the outcome of a study (Hill
Variability is minimized by standardizing the delivery of health care within a study, and laying this down in detail in a study protocol (Chalmers 2001). The study protocol can specify, amongst others, dosages, time schedules, administration routes, the characteristics of patients, additional care, medication adherence measures, the health care setting, treatment monitoring, and safety measures. Health care is subsequently to be delivered by strictly following this protocol.

However, it has long been observed that interventions often perform worse in regular clinical practice than in the setting of a study with standardized care, for instance by having a lower efficacy or more serious adverse effects. This problem is known as the efficacy–effectiveness gap (Eichler et al. 2011). One of the causes of the efficacy-effectiveness gap is the fact that patients in trials differ from patients in regular clinical practice with regard to e.g. genetic makeup, physiology, comorbidity, disease burden, stress, physical activity and diet. This issue has been discussed in the previous section. Furthermore, the efficacy-effectiveness gap depends on differences in how health care is delivered in trials and in regular care. Examples of these differences are: more inappropriate or off-label prescribing in regular care, more co-prescribing with an interacting medicine, more continued prescribing to non-responders, more medication errors, poorer adherence to prescribed treatment regimen, more treatment discontinuation, more taking of ‘drug holidays’ and more inadvertent overdosing in regular care (Eichler et al. 2011).

Although this efficacy-effectiveness gap can cause problems for all sorts of patients, I suggest that it may affect disadvantaged patients to a higher degree because the care they receive in clinical practice is likely to be farther removed from the standardized care in trials. There are two possible reasons why this may be so: the first has to do with a patient’s capabilities and the second with proper access to care. Firstly, people with little resources or socially deprived people with little social support structures may in some cases be less capable of adapting their life to an intervention in the exact way it has been studied in an RCT. This idea is supported by the fact that patients who have a low level of education, low self-efficacy expectations, a high degree of life stress, or low level of health literacy experience greater difficulty in adhering to prescribed medicine regimens in general and are more prone to discontinue their medication or take a ‘drug holiday’ (Christensen 2004). In the specific case of
hypertension management, research has shown that problems with adherence in ethnic minorities is a main cause for the fact that antihypertensive medications are less effective in these groups (Holmes et al. 2012).

The second reason why for the disadvantaged regular clinical care can differ so much from standardized care in trial settings, is that people with little resources, those suffering from discriminatory treatment, and other disadvantaged groups can have more difficulties in getting adequate support from the healthcare system (Epstein 2007). For example, disadvantaged patients are more likely to be subject to inappropriate or inadequate prescribing, medication errors and inadvertent overdosing (Epstein 2007). A lack of proper support from the healthcare system makes it harder to deliver the intervention in the exact same way that it was studied in the trial setting, widening the efficacy-effectiveness gap.

To close the efficacy-effectiveness gap so-called ‘pragmatic clinical trials’ have been proposed: trials in which care is as similar as possible to regular clinical care. Pragmatic trials study patients that are similar to the regular patient population; this has been discussed in section 2. Furthermore, pragmatic trials strive to make care as it is delivered within a study more consistent with regular clinical care, i.e. they make care within a trial more variable and less standardized (Tunis SR 2003, Ware and Hamel 2011). Although the internal validity of pragmatic trials can be problematic and although they require larger sample sizes (Ware and Hamel 2011), their results indeed better apply to the average patient, which means they are fairer towards disadvantaged patients. However, following the same line of reasoning as in section 2: although pragmatic trials are fairer to the average patient, their results still do not apply properly to patients who are non-average and who have the most difficulties with adapting to a treatment or with getting adequate support from the healthcare system.

In sum, in this section I have argued that RCTs focus on standardized health care interventions, which can result in an efficacy-effectiveness gap. This gap, which cannot be eliminated completely for all patient groups by using a pragmatic trial approach, may affect disadvantaged patients more, amounting to an unfair distribution of the benefits of research results.
6.4 Standardized interventions

In the previous section I have argued that standardized care (the way in which interventions are used) in RCTs can lead to unfairness. In this section I argue that RCTs are more suited to study the kind of interventions that are, or can be, standardized, and that this can be unfair to disadvantaged patients. In particular, I argue that RCTs are generally more suited for studying medicines, which can often easily be standardized, than for other types of interventions such as lifestyle interventions, which can be difficult to standardize. Not properly studying these types of interventions is unfair to disadvantaged patients because they often experience the most problems living a healthy life. For example: socio-economically disadvantaged groups and ethnic minorities more often have unhealthy lifestyles in terms of cardiovascular risk factors like obesity, physical inactivity and unhealthy diet. So, these patient groups could especially benefit from proper medical research on interventions that target these factors.

RCTs require that interventions are, or can be, standardized by requiring that interventions are of constant quality and reduced to their essential element(s). RCTs require that interventions are of constant quality because this helps to reduce random error (noise) in a study and to arrive at clear results (a clear-cut differences in outcome between groups with different interventions). RCTs also require that interventions are reduced to their essential element(s), because reducing the number of components in an intervention helps to eliminate sources of variability. This means that there is a drive in RCTs towards subdividing interventions into their underlying components and to study these components in isolation.

Besides methodological reasons for requiring standardized interventions, this requirement also has its roots in the (historical) link between RCTs and the regulation of medicines. The development and use of RCTs has from the outset been strongly tied to the need to evaluate medicines and regulate their market access, and information about quality and quality assurance measures have been important considerations for regulatory authorities (Streptomycin in Tuberculosis Trials Committee 1948, The Council of the European Economic Community 1965). Regulatory authorities only allow a medicine to enter the market if it is clearly defined and of constant quality: the
composition of a product should be described in detail and the composition of each specimen should be identical and free of any contaminations.

Due to the historical link between RCTs and medicines, it will come as no surprise that most medicines can meet the requirement for standardization. Medicines are generally industrially manufactured on a large scale, which makes it possible to define their composition in detail, to make products of a constant quality, and to purify products so that they contain just a single active ingredient. However, other types of intervention, such as psychological or social intervention (e.g., psychotherapy, social psychiatric interventions, motivational therapy for substance abuse, lifestyle interventions, occupational health interventions, (nursing) home care and the facilitation of informal care), can have more difficulties in meeting the requirement for standardization.

Here, I take lifestyle interventions for hypertension management as an example. Some lifestyle interventions, such as diet and exercise, are by their nature less clearly defined and of less constant quality because they depend heavily on the activities performed by individual health care providers and patients in daily practice. For instance, although it may seem straightforward to advise a patient to eat a certain amount of vegetables and fruit per day, in practice patients will vary considerably in what kind of vegetables and fruit they eat and how they prepare them (raw, cooked, baked etc.). This variability makes this kind of intervention less suitable to be studied with an RCT because it would be hard to get a clear difference in outcome between groups, and to link the outcome to the intervention. Further standardizing a patient’s diet by prescribing some sort of standard diet would not be an adequate solution: it may work for a short while within a trial but is clearly not a realistic option for daily life. Also an apparently simple advice to walk an hour a day will in practice be taken up in different ways: how far and how fast should a patient walk; does going shopping count or should it be just walking; are breaks allowed and for how long, etc.?

A further problem of studying lifestyle interventions in RCTs is that blinding of health care providers and patients, a means of standardizing interventions, is near impossible. Again, this will make treatment groups more variable and possibly biased, leading to less clear differences in outcomes of groups. Moreover, one of the strengths of lifestyle interventions actually is their variable nature and the fact that they can be
tailor-made based on the skills of the health care provider, the health care setting, the characteristics of a patient, a patient’s preferences and his/her current situation. For example, advice to walk an hour a day may be suitable for someone living next to a nice park, but may be less appropriate for someone living in a bad neighborhood; in such a case, it would be better to give a tailor made advice and suggest going to a gym. Standardizing such interventions in the context of an RCT would make them appear less effective than they in reality are.

A final problem of studying lifestyle interventions in RCTs is that they often cannot be reduced to their essential element(s) because they work by virtue of being part of a holistic approach: they combine a number of elements or consist of several interlinked steps. For example, whether a patient’s diet will contain little saturated fatty acids depends on the overall diet and not just on reducing fatty acids by following an advice to eat low fat cheese, especially if it causes a patient to eat more other high-fat products. A further example related to hypertension management is how to educate and stimulate patients to engage in physical activity (simply giving advice is often not enough). Current approaches propose stepwise learning to increase motivation and suggest to address the complete range of individual and contextual factors that can inhibit or enable physical activity (Stokols 1996, Whitlock et al. 2002). An individual’s socio-economic circumstances, cultural background and living environment are such factors. Splitting up such an intervention by only addressing a single enabling factor or focus on just one step in a learning cycle will make them less effective. So again, the point is that standardizing lifestyle interventions in order to study them with an RCT will make them appear less effective.

In sum, medicines have set the standard for interventions in RCTs: interventions should be clearly defined and of constant quality, and reduced to their essence. I have argued that this is a standard that particular types of interventions, such as lifestyle interventions, can have difficulties to meet, which is unfair to disadvantaged patients because they often experience the most problems living a healthy life. Moreover, I have argued that standardizing such interventions in order to study them with an RCT is also not an adequate solution as it can render them less effective. So, insisting on RCTs as the gold-standard methodology for studying lifestyle interventions
can deprive disadvantaged patient groups of proper knowledge on relevant interventions.

6.5 Discussion

In the three previous sections I discussed how RCTs reduce variability in a study by means of standardization. Standardization aims to prevent a possible distortion of the most important measure in RCTs, the effect of the ‘new’ intervention as compared to the control/old intervention (Chalmers 2001). I argued that RCTs strive to reduce variability between intervention and control groups by standardizing (or ‘niche-standardizing’) patients. Furthermore, I argued that RCTs try to prevent variability in how interventions are used in a study through standardization of care with strict study protocols. Finally, I argued that RCTs require that interventions themselves are standardized: clearly defined, of constant quality and reduced to their essence.

A weakness of my study is that it has an exploratory character and is for a large part based on inferences from ‘circumstantial’ evidence. Especially the link between standardization in research methods and unfairness deserves further study, for instance by conducting in-depth case studies of how individual RCTs have affected particular disadvantaged patient groups. Furthermore, although little direct evidence was available to support the link between RCTs and standardization, this link appears to be somewhat more robust because standardization in RCTs is not an isolated phenomenon, but can be seen as healthcare’s version of how variables are often controlled in scientific experiments: by creating an artificial world, the ‘laboratory’, in which variation is reduced by standardization (Latour 1987). Furthermore, standardization in medical research should also be viewed in light of broader efforts to standardize healthcare (Timmermans and Berg 2003).

The reduction of variability through standardization, as it is done in RCTs, has several important advantages. For instance, it can help to reduce systematic error. This will improve the internal validity of research, meaning that the outcomes of a study can justifiably be ascribed to the effects of the interventions. Reducing variability through standardization also helps to reduce noise (random error). This increases the precision
of measurements and results in clear outcomes of a study. A final benefit of standardization is that it helps make the results of studies suitable for comparison and aggregation, e.g. in meta-analyses.

However, although standardization in RCTs helps generate valid results, this does not necessarily mean that those results are relevant to all patient groups. In the previous three sections I argued that the methodology of RCTs can lead to an unfair distribution of the benefits of research results: (1) research results in standardized study groups can be difficult to extrapolate to disadvantaged groups in clinical practice; (2) the focus on standardized health care can lead to an efficacy-effectiveness gap, which can affect disadvantaged patients more; and (3) the focus on clear and simple interventions can deprive disadvantaged patient groups of proper knowledge on more complex interventions, such as lifestyle interventions.

The first two problems can partially be addressed in clinical practice by health care providers. Given a fair amount of clinical expertise, it is possible to extrapolate from knowledge on standard patients and standardized care to a particular patient in regular clinical practice, and adjust the therapy accordingly. The second problem can additionally be addressed by making sure that disadvantaged patients get appropriate support from the healthcare system and by providing them with complementary care to help them adapt to interventions, for instance by stimulating adherence. However, when there is no adequate evidence available at all concerning a (lifestyle) intervention (the third kind of problem), the health care provider is left solely with his clinical judgment to support a therapeutic approach.

Besides addressing the problems of RCTs in clinical practice, a more direct solution would be to address the problems of the methodology of RCTs themselves. As I described in the previous sections, several developments already focus on reconsidering where, to what extent and how to standardize in RCTs. However, I also pointed out that these solutions can have their own methodological problems and do not always lead to adequately addressing the needs of disadvantaged groups.

An alternative solution to lowering the degree of standardization in RCTs would be to shift the focus of medical research more towards methodologies in which standardization is a less central concept, for instance, by focusing more on observational studies. Observational studies study phenomena in settings closer to real life, as opposed
to the laboratory-like settings of RCTs. To what extent, and in what ways other methodologies can be a proper alternative to RCTs is currently debated with regard to making the marketing authorization of new medicines more adaptive (Eichler et al. 2012). It has been proposed that for some types of medicines it should be possible to grant initial marketing authorization on the basis of smaller RCTs with less statistical power, and use observational methods to generate additional knowledge in a later stage (De Jong et al. 2013, Eichler et al. 2012). Such an approach could lead to more relevant knowledge on a medicine for disadvantaged groups. Although this debate is fueled by concerns about innovation and costs in medicine development and not by a concern for justice, it does signify a shift away from the prevailing position that the RCT is the gold standard.

Furthermore, I suggest that it should be considered on a case-by-case basis whether RCTs will result in relevant knowledge for all patient groups, including the disadvantaged ones, or that other methods might be more appropriate. For example, in the previous section I have argued that for the study of interventions that are complex and difficult to standardize, such as some lifestyle interventions, RCTs are not a suitable methodology. I conclude by drawing attention to a particular kind of observational methodology – qualitative methods – that could complement RCTs and could help to generate more relevant knowledge for disadvantaged groups.

The main rationale for highlighting qualitative methods as a method that could do more justice to disadvantaged patient groups is that they strive for the opposite of standardization: we could call this ‘adaptation’. As discussed, RCTs strive for standardization in order to control and reduce variability. They do so in a laboratory-like setting with the goal of a precise measurement of one component of the object of study. The philosophy of qualitative methods is a mirror image of this: a vital aspect of qualitative methods is that in order to get a comprehensive understanding of the how and why of phenomena, qualitative methods specifically search and describe the variability of real life settings. They do so by adapting methods to both the topic of study and during the study also (Taylor et al. 2010). Qualitative methods are more concerned with generating hypotheses than with testing them and more with revealing new kinds of distinctions than with generating evidence on a distinction provided by the researcher (Pope and Mays 2006). As qualitative methods aim to study variation and
distinctions, I believe that this implies that instead of studying standard patients in
standard settings, receiving standardized therapy, qualitative studies can look for
evidence on non-standard patients in non-standard settings, receiving non-standardized
therapy. Thereby, qualitative studies may result in knowledge that is more relevant to
disadvantaged groups. Actually, many qualitative studies focus explicitly on cultural and
social differences. Below, I give some examples of how a qualitative approach could
relate to patients, care and interventions in medical research, and I indicate what
bearing this will have on justice.

Firstly, in order to gain insight into what kind of patient groups are relevant for
studying a disease and its treatment in RCTs, one may first have to generate hypotheses
about how people can be grouped, and how characteristics of a group can interfere with
the effects of an intervention (Epstein 2007). Qualitative methods can help to generate
these hypotheses and make them reach beyond the ‘usual suspects’ defined by sex, age
and ethnicity. For example, it is known that ethnicity is a determinant of how patients
explain and manage hypertension (Morgan 1995). However, this does not automatically
mean that belonging to an ethnic group is the actual causal factor in this. For example, a
qualitative study showed that how patients manage hypertension might be less related
to ethnicity itself than to a history of migration (Beune et al. 2006). This suggests that
migrants might be a relevant subgroup to study in RCTs on hypertensive medication.

Furthermore, to close the efficacy-effectiveness gap one should not just make
care in trials more lifelike, but one should study the gap itself and how it is related to
various patient groups. Qualitative methods could help to understand which groups
experience difficulty in adapting to interventions and for studying the underlying
mechanisms. For instance, a qualitative study showed that poor medication adherence
in Surinamese and Ghanaian patients was related to patient worries about negative
effects on sexual performance and that they discontinued medication when visiting
their homelands (Beune et al. 2008). This knowledge could be used in clinical practice
to tailor information to these groups in order to improve adherence.

Finally, qualitative methods appear to be a viable alternative for studying the
more complex psychological and social interventions, such as some lifestyle
interventions, that are especially needed by disadvantaged patient groups. Qualitative
methods can help to study the main strength of these interventions: how several
components are combined into a successful therapy (Mol 2006). Furthermore, they can be used to study how health care professionals exercise clinical judgment and tailor therapies to patients (Lutfey 2005, Mol et al. 2010). For example, helping patients to engage in physical activity requires insight into what relevant enablers or barriers are. For instance, a qualitative study showed that specific barriers for Ghanaian and Surinamese patients are preferences for large body sizes and unfamiliarity with recommended physical activities such as cycling (Beune et al. 2010). Interventions could be adapted accordingly: focus less on weight loss and more on physical activity, and promote walking instead of cycling.

In sum, RCTs tend to standardize patients, care and interventions, meaning that in some cases less relevant medical knowledge will become available for disadvantaged groups. I do not conclude that RCTs should not be used for studying health care interventions, but suggest that RCTs should be considered less of a gold standard. To put the current status of RCTs in perspective, I quote what one of the major contributors to the methodology of clinical trials, Bradford Hill, said in 1952: “The statistically guided therapeutic trial is not the only means of investigation and experiment, nor indeed is it invariably the best way of advancing knowledge of therapeutics. I commend it to you as one way” (Hill 1952). I believe that the benefits that derive from research results could be more fairly distributed if RCTs would standardize less and would be complemented more often by observational, and especially qualitative, methods. This could help to generate knowledge that is more relevant for disadvantaged patient groups and so to better do justice to the multitude of differences between patients. These are considerations that could be taken into account in ethical oversight.
References


7

DISCUSSION

(Dis)approving and improving in action

7.1 Introduction

This thesis deals with ethical oversight on medical research with human subjects. In the prologue to this thesis I recounted the story of Jan. Jan was suffering from prostate cancer and chose to participate in a scientific study in order to help improve the treatment of this disease. Unfortunately, the study led to severe side effects for Jan. Although Jan did not feel that his physician or the researchers were culpable for his situation, he was left with many questions. Had it been a good idea to allow an alteration of his treatment based on the MRI? Was he informed properly about the risks or could additional information have changed his mind? Was the knowledge gained by this study worth the risks to him and other patients? And he also wondered: Who decides about all of these issues? Who ensures that these things are in order? And how do they do that?

In most countries a system for ethical oversight on medical research has been in place for many years. This system aims to ensure that the ethical quality of research is in order and that subjects are protected. Research Ethics Committees are considered the key components in this system. However, the practice of ethical oversight is not transparent, but rather a ‘black box’: we know little about what is going on inside the system. The central aim of this thesis is to open this black box and to help understand how ethical oversight, especially by Research Ethics Committees, works in practice. A secondary aim of this thesis is to help improve the functioning of ethical oversight. In the previous chapters, I have therefore studied various aspects of the practice of ethical
oversight in The Netherlands and the U.S., focusing especially on the work of Research Ethics Committees.

In this concluding chapter, I work towards a deeper understanding of ethical oversight. Based on a synthesis of the contrasts uncovered in the studies in this thesis, I argue that there are two philosophies for doing good in ethical oversight: '(dis)approving' and 'improving'. Subsequently, I reflect on the research methods used in this thesis and I discuss the strengths and weaknesses of both approaches to oversight. Finally, I discuss how the two approaches to oversight interact, and how they can best be combined.

7.2 Synthesizing a deeper understanding of ethical oversight

So what understanding of ethical oversight emerges from my studies? As explained in the introductory chapter, I chose to study ethical oversight in practice and focused on contrasts – oppositions that emphasize differences – as a way of getting into the 'black box' of oversight. In chapter two and three this led me to mark out two contrasting ways of doing ethical oversight: in chapter two I distinguished working within a repertoire of rules from working within a repertoire of production, and in chapter three I distinguished compliance monitoring from quality improvement monitoring. In chapter six I discussed the ethical implication of two philosophies behind research methods: I contrasted the philosophy of standardization in randomized controlled trials with the philosophy of adaptation in qualitative research. Furthermore, in chapter four I compared published with non-published studies, and I discussed two different ways of dealing with non-publication: disapprove a proposal or discuss concerns with investigators. Finally, although chapter five dealt predominantly with oversight on tissue storage and not on research, looking back on this study we can see a tension between approaching this issue from the perspective of principles available in the ethical and legal literature and from the perspective of the practical aspects of the actual storage of tissue.

I argue here that these contrasts signify a more general contrast in oversight. Analyzing my findings in conjunction shows that they fit into two philosophies for
doing good in ethical oversight: ‘(dis)approving’ and ‘improving’. I distinguish these two approaches to oversight on the basis of how they differ on four aspects: their aim, the sort of activities; the kinds of relationships; and the kinds of objects (Table 1).

The (dis)approving approach works towards (dis)approval of research by ensuring that research is ethical, is carried out within relationships based on authority, and focuses on documents. The improving approach on the other hand, works towards improvement of the ethical quality of research, is carried out within relationships based on equality, and focuses on actual practices, i.e. instead of on the paperwork that describes a physical reality, on the physical reality itself.

Before elaborating on how a (dis)approving and an improving approach to ethical oversight differ, I first want to stress that their higher goals are identical. Both approaches are concerned with the ethical quality of research: they strive to protect the interests of human subjects in medical research while allowing scientific progress to help future patients. It is also important to note that the distinction between the two approaches to ethical oversight is of an analytical nature: in practice, their differences will be a matter of degree and not clear-cut. For example, in reality, a (dis)approving approach will not necessarily be restricted to documents but will also be concerned with actual research practices. However, my point is that a (dis)approving approach will focus relatively more on documents and less on the actual research practices. Conversely, an improving approach will not necessarily be restricted to actual research practices but will also be concerned with documents. However, an improving approach will focus relatively more on actual research practices and less on documents. This

Table 1: Two philosophies for ethical oversight – (Dis)approving and Improving

<table>
<thead>
<tr>
<th></th>
<th>(Dis)approving</th>
<th>Improving</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aim</strong></td>
<td>(Dis)approval</td>
<td>Improvement</td>
</tr>
<tr>
<td><strong>Activities</strong></td>
<td>Ensuring</td>
<td>Advising</td>
</tr>
<tr>
<td><strong>Relationships</strong></td>
<td>Authority</td>
<td>Equality</td>
</tr>
<tr>
<td><strong>Objects</strong></td>
<td>Documents</td>
<td>Practice</td>
</tr>
</tbody>
</table>
notwithstanding, to get a clear picture of the (dis)approving and improving approaches, I will focus on their contrasts in the next two sections.

7.3 (Dis)approving medical research

The (dis)approval approach to ethical oversight is defined by the fact that its main activity consists of ‘ensuring’ that research is ethical and that it aims to come to an ‘approval’ or ‘disapproval’ of research. Ensuring means that parties work on safeguarding the ethical quality of research, making certain that research is ethical by setting up some sort of external controls. Approval or disapproval is the official endpoint of this process, the guarantee that research is ethical. For example, chapter two showed that Research Ethics Committees (working within a repertoire of rules) ensure that research is ethical by applying rules and regulations, the legitimacy of which is ensured by democratic processes. Furthermore, reviewing proposals in the repertoire of rules results in a final and official approval (or disapproval) of research proposals. This gives patients certainty that research is safe and controlled by independent authorities. In chapter three we saw how research institutions have set up compliance monitoring programs, by some called the “IRB police”, to ensure that researchers comply with rules and the approved proposal when research is ongoing. And finally, in chapter six we described the activity of ensuring with regard to research methodologies. Randomized controlled trials aim to ensure that a study provides a clear and definitive answer by standardizing the study approach and the execution of the study.

A further defining aspect of the (dis)approval approach to ethical oversight is that relationships are based on ‘authority’. The highest levels of authority are the law and governmental oversight bodies, and below that there is the authority of Research Ethics Committees over research and of the research institution over its researchers. These authoritative relationships help to ensure that research is ethical. Chapters two and three, for example, showed that the relationship between researchers and parties involved in oversight on ongoing studies, Research Ethics Committees or other offices at research institutions, is based on authority. Researchers ‘must’ clarify, motivate or
change their research proposals, and in compliance monitoring programs, researchers can be disciplined and obliged to take corrective actions. I also discussed the hierarchical relationship between Research Ethics Committees and researchers in chapter four, where I discussed whether a committee should withhold approval based on the likelihood of non-publication, or could even sanction non-publication.

The final aspect of the (dis)approval approach to ethical oversight is that the main objects of this type of oversight are ‘documents’ that describe research and prescribe research oversight, or in other words: ‘paperwork’. Documentation is used as a means to ensure that research is ethical. For example, documents (i.e. regulations) grant committees the authority to oversee research and prescribe how the ethical quality of research should be ensured. Committees in their turn use documents (i.e. letters) to transfer this authority to research proposals; the approval letter ensures that a study is ethical. Documents are also used for (legal) accountability: committees should be able to show proper documentation of their decisions to the outside world, and researchers should be able to show that their research is ethical by having the approval documents in place. This is not to say that a (dis)approval approach cannot, for example, focus on the practice of ongoing research, but does mean that in doing so, it will focus predominantly on research documents and not on the actual practices and physical reality of research. This point is illustrated by chapter three, where I describe how ‘compliance monitoring’ programs, which aim to oversee the actual conduct of research, focus solely on the documentation of research. Furthermore, chapter two described how the input for ethical review by the Research Ethics Committee consists of documents: rules, regulations, forms, research proposals and letters from researchers. The output of the review process also consists of documents: a letter to the researcher with questions and/or a (dis)approval decision. Furthermore, in chapter three I discussed how compliance-type monitoring will result in reports that describe the quality of a study, and how these will lead to follow-up reports by researchers. A final illustration of the focus on documents can be found in chapter five. Although this chapter was not primarily concerned with the ethics of research conduct, but rather with the ethics of dealing with objects of clinical research (in this case residual tumor tissue), here too, ethical problems are dealt with by focusing on documentation: creating a guideline, based on scientific literature.
7.4 Improving medical research

The improving approach to ethical oversight is defined by the fact that its main activity consists of ‘giving advice’ on research and that it aims to ‘improve’ the ethical quality of research. Advice consists of recommendations on how to improve the information to subjects, how to decrease burdens and risks, how to improve the research methodology, and how regulatory requirements could be met. For example, in chapter two I described that Research Ethics Committee members do not consider research proposals as a given to which they can only say ‘yes’ or ‘no’, i.e. approve or disapprove. The committee will give researchers advice on how to better protect the interests of subjects, and how to make proposals fit regulatory requirements or the local research setting. This approach helps make disapproval of research proposals very rare. I encountered the same attitude in chapter three: monitors involved in quality improvement programs worked towards improvement of the actual conduct of research. Furthermore, the attitude of giving advice also works the other way around: quality improvement monitors can collect information on the needs and wishes of researchers, and use this to help improve the arrangements and facilities for research at the institution.

A further important aspect of the improving approach to ethical oversight is that relationships are based on ‘equality’. Researchers and parties involved in oversight, such as Research Ethics Committees, work together on a more or less equal footing at improving the ethical quality of research. Communication between parties takes the form of advice. In chapter two I discussed the relationship between Research Ethics Committee and researchers and concluded that they were (also) co-workers, both parties working on the ethical quality of research in a collegial fashion. The committee gives advice to researchers, enters in discussions with researchers and can also make use of informal forms of communication because committee members are in close contact with researchers. Chapter four described an example of how a Research Ethics Committee can discuss ethical issues on an equal basis: I suggested that committees could use information on probable non-publication for discussing their worries with researchers. As a final example: in chapter three I described that quality improvement monitoring can only be done properly if monitors work together with researchers in an
atmosphere based on trust. This is a relationship between equals, both parties giving each other feedback on how to improve research and research oversight.

The final defining aspect of the improving approach to ethical oversight is that the main objects of this type of oversight are ‘practice’. Although paperwork can also be seen as being part of a practice, the point here is that the improving-type of oversight strives to go beyond paperwork and focuses on the actual practices and physical reality of research. For example, I described in chapter two how a committee can work within a repertoire of production, busying themselves with the practical consequences of oversight, down to a concern for the time and money involved in research and oversight. Furthermore, the committees will take the time to review research proposals in multiple rounds, delving into the practical consequences of ethical considerations in order to give tailor-made advice. Another illustration of a concern with actual practice can be found in chapter five. The issue of proper storage and use of residual (tumor) tissue was put on the agenda, inspired by an actual case from clinical practice. Furthermore, the analysis of this case was fed by knowledge about the practicalities of managing tissue storage. And finally, although the quality improvement programs described in chapter three looked at paperwork (research records) and resulted in paperwork (quality improvement reports), they also went beyond paperwork and focused on practice: monitors would see to the actual conduct of research and discuss their findings in person with researchers, leading to practical advice.

7.5 Reflection on the research methods

I have used several methods in this thesis to answer my research questions and have discussed their specific strengths and weaknesses in the corresponding chapters. Here I reflect on the approach of this thesis as a whole.

To study ethical oversight on medical research with human subjects in practice, I studied Research Ethics Committees because they have a central position in the system for ethical oversight, but I also collected data regarding other offices at research institutions responsible for oversight. My data came from several U.S. and Dutch research institutions and Research Ethics Committees. Much of the in-depth data from
The Netherlands came from one large academic medical center and its corresponding Research Ethics Committee. My thesis is thus based on data from a limited number of settings. However, by collecting data from various sources, ranging from literature databases to ethnographic studies, I was able to get a rich understanding of those settings. Furthermore, by using a multi-method approach in this thesis, I could compare ('triangulate') the outcomes of different studies and test the internal validity of my findings. Comparing the findings of previous chapters to the two types of oversight presented in this concluding chapter, I have not found important inconsistencies. The validity of my findings was further supported by discussing them with people who worked in the settings I studied.

However, in qualitative research, where the generation of data through observation and the subsequent interpretation of data depends heavily on the individual researcher, the researcher can be an important source of distortion of the outcome of a study. It is therefore important to reflect on the assumptions and initial hypotheses I brought into this thesis. My initial assumption about Research Ethics Committees was that their core activity consisted of discussing how to support a decision to (dis)approve a research proposal, and that members were solving differences of opinion by justifying their opinions to each other. Taking these discussions and difference of opinion as a focus point also fitted well with my methodological approach of looking for contrasts. This interest in how people justify decisions and opinions led me to the book 'On Justification, Economies of Worth' which argues that people justify their actions to others within six different 'worlds' (Boltanski and Thévenot 2006). This idea of a moral pluralism fitted well with my intuition, inspired by a thesis on public debate about ethical issues in medicine (Trappenburg 1993), that Research Ethics Committee discussions consisted of some sort of pluralism (or more precisely, a dualism). My earliest conception of this duality in ethical oversight derived from interviews with people involved in monitoring ongoing research during a visiting scholarship at New York University. I hypothesized then that there were two ways to perform ethical oversight, 'justice' and 'education'. This hypothesis fueled later studies and evolved finally into the distinction between '(dis)approving' and 'improving'. This insight challenged my initial assumption that Research Ethics Committees are predominantly involved in justificatory discussions in order to (dis)approve research proposals. I had to
conclude that, yes, one of their activities focuses on ‘(dis)approval’, but there is also another way of doing oversight, ‘improving’.

I have discussed the generalizability of my findings on specific aspects of ethical oversight in each corresponding chapter, so here I focus on the question whether the overall distinction between \textit{(dis)approving} and \textit{improving} might apply outside the settings I studied. My finding that Research Ethics Committees use two approaches to oversight can be explained by the fact that they are positioned between two worlds: on the one hand they are part of the judicial world of government regulation, in which approval and disapproval of research is of legal significance, and on the other they are part of the world of the research industry (this includes research institutions for medical research). Since many Research Ethics Committees are similarly positioned, I believe they might use these two approaches to oversight too.

Another reason why I think my findings might apply more broadly, is that many Research Ethics Committees are alike with respect to the kind of research proposals they review, their composition, their legal status, and the process of ethical review (Centrale Commissie Mensgebonden Onderzoek 2007, Centrale Commissie Mensgebonden Onderzoek 2010, De Jong et al. 2010, Decullier et al. 2005, Dickersin et al. 1992, Easterbrook et al. 1991, Easterbrook and Matthews 1992, Fitzgerald et al. 2006, Hall et al. 2007, Stern and Simes 1997). On the other hand, committees also differ. Committees are known to differ with respect to their decisions about particular research proposals, their workload, and the degree to which they are organized, e.g. in terms of working procedures and office support (Abbott and Grady 2011, Centrale Commissie Mensgebonden Onderzoek 2010). For example, my findings on oversight on ongoing research should be generalized with caution, because I found considerable differences between monitoring programs, especially with regard to the position of Research Ethics Committees.

Previous research provides some further confidence that the distinction between \textit{(dis)approving} and \textit{improving} is valid and could apply more broadly. Noah, for example, has discussed whether U.S. Research Ethics Committees are more like adjudicatory bodies making final, legal decisions on research proposals, or more like peer review mechanisms providing iterative feedback to improve research proposals (Koski 2007, Noah 2004, Weijer et al. 1995). Other authors have distinguished a
compliance philosophy from a quality assurance philosophy (Dyck and Allen 2013, Koski 2007, Weijer et al. 1995). Dyck and Allen, for example, have recently discussed the role of Research Ethics Committees and argued that instead of promoting compliance with inflexible and universal rules, the role of a committee should be to facilitate and resource the reflective practice of researchers. They argue that a simple, but significant, shift would be to move ethical review from approving a proposed project to providing guidance and feedback on submitted projects (Dyck and Allen 2013). Also, scholars in the field of ethical oversight supported the validity of (dis)approving and improving in conferences and interviews where I presented my findings. Moreover, my conceptualization of (dis)approving versus improving fits a much broader sociological distinction between ‘civic’ and ‘industrial’ worlds (Boltanski and Thévenot 2006). Still, I have studied only a limited number of settings and I did not investigate other parties involved in ethical oversight besides Research Ethics Committees and research institutions, such as governmental organizations, sponsors and research leaders. Therefore, prospective testing of my analysis of ethical oversight is needed to see whether (dis)approving versus improving captures a clear distinction and to what extent it applies to other settings.

7.6 (Dis)approving and improving in action

In this section I discuss several strengths and weaknesses of both approaches to ethical oversight. In the ensuing section I discuss how the two approaches to oversight interact: how they interfere with each other and how they can be combined.

A major strength of the (dis)approving approach is that it aims to ensure that every research proposal has undergone an ethical evaluation and that the same rules and standards are applied in each case. This can help to protect research subjects from excesses in medical research and can also be beneficial to researchers: they know what to expect from research oversight and can see to it that their research complies with ethical standards. On the down side, the effort involved in complying to detailed rules and standards can slow down scientific progress, and the rigidity of regulatory requirements can impede research with regard to special situations (e.g. emergency

The fact that relationships in the (dis)approval approach are based on authority can be both a strength and a weakness. A strength of this kind of relationship is that it allows oversight bodies to prevent, stop, and correct research that does not conform to ethical standards, and allows for disciplining researchers who have been involved in severe research misconduct. However, oversight that is perceived as policing can hamper public trust, and the use of punitive measures can destroy the ‘informal monitoring system’, i.e. people helping voluntarily with oversight by identifying problems (Holmberg 2004, Levine 1980, Weijer et al. 1995). Finally, relegating the question of whether research conforms to ethical standards to an authority (the Research Ethics Committee) may cause researchers to feel less responsible for the protection of human subjects.

The focus of the (dis)approval approach on documents (e.g. research proposals, approvals, and informed consent forms) can help make explicit to patients, researchers, committees, institutions and other oversight bodies what a particular study entails. This can help ensure that ethical standards are met. However, having detailed documents in place can lead to the erroneous belief that this automatically leads to an adequate protection of the interests of subjects: e.g. detailed informed consent forms are notoriously difficult to understand (Flory et al. 2008). Furthermore, relying on documents can lead to a false sense of security: it is not evident that the paperwork of research (and research oversight) corresponds to the actual practice of research (and research oversight), and thus whether the interests of subjects and science are adequately protected. A benefit of a focus on documentation is that it helps to efficiently transfer research and oversight from one setting to the next, for instance in the case of multi-center studies. And finally, having research and oversight decisions on record protects researchers and research institutions from legal liability and interference by other, ‘higher’ oversight bodies.

Having discussed the strengths and weaknesses of (dis)approving, I now turn to improving. A major strength of improving is that it aims to give researchers practical advice on how to improve their proposals with respect to ethical standards, regulatory requirements and the local research setting. The flexibility of the improvement
approach helps to make advice tailor-made and to find solutions for specific, idiosyncratic problems. *Improving* can also take the form of educational activities, which can help to raise the ethical quality across the entire research community. Furthermore, *improving* works both ways: information from researchers can be used as feedback to improve the quality of the ethical oversight system and the institution’s arrangements for research. However, the flexible nature of *improving* could become a weakness if Research Ethics Committees would be too flexible in how they review research proposals. It could mean that not every research proposal would undergo a proper ethical evaluation; potentially leading to research that does not meet ethical standards. Furthermore, a flexible approach to oversight can make the terms of reference for researchers unclear, which could make it difficult for them to set up their study in such a way that it will be acceptable to all parties. This could especially be problematic for multicenter studies: needing to negotiate proposals for research anew with every committee would grind these studies to a halt.

Furthermore, carrying out oversight within relationships based on equality can be beneficial to both the quality of research oversight and of research, because it stimulates an unrestricted exchange of information between oversight parties and researchers (McCormack et al. 2012). Working within equal relationships supports a constructive discussion between researchers and oversight parties and allows for additional forms of communication outside (formal) meetings and letters. Furthermore, the fact that researchers and oversight parties are on a somewhat collegial footing supports an atmosphere of trust in which an ‘informal monitoring system’ can prosper and can make researchers feel more responsible for the protection of human research subjects (Levine 1980). A major drawback of performing ethical oversight within an equal relationship is that it can be difficult to take corrective actions towards research or researchers when ethical standards are breached.

The improvement approach to ethical oversight focuses on the actual practice of research, which is a major strength. Overseeing the actual conduct of research and how it affects human subjects in practice can lead to practical and tailor-made advice for both researchers and the research institution on how to improve subjects’ protection. The focus on practical aspects of oversight also helps to take into account how oversight is implemented, for example to what efforts and costs oversight amounts,
which can help to make oversight less burdensome to researchers. A weakness of a focus on practice instead of on documents is that there will be less documented information available about research and its oversight, making the system less transparent to outsiders and making it more difficult to exchange information from one research setting to the next (e.g. in multi-center studies). A final weakness of a focus on practice is that if researchers, committees or other parties engage with the law (e.g. governmental oversight bodies or liability cases), the fact that things were in order in practice is not sufficient: the legal system looks at documentation (records), and when documentation is inadequate, parties are vulnerable to legal action.

7.7 (Dis)approving and improving in interaction

Both philosophies for ethical oversight thus have their specific strengths and weaknesses for research subjects, researchers and oversight bodies, so choosing between them is complicated. Moreover, the choice will depend on the specific needs and wishes of a research institution or Research Ethics Committee, on its organizational structure, on the type of research conducted there, and on the research culture. These intricacies make it impossible to give a general advice about which type of oversight to choose. Moreover, little is known about the actual effects of ethical oversight on research subjects’ protection and research, let alone about the effects of different approaches to oversight (Abbott and Grady 2011, Coleman and Bouesseau 2008). This too, makes choosing between them difficult. Based on my findings, I think that the following general advice is appropriate: be aware that it is possible to choose between two approaches to oversight and take heed of their specific strengths and weaknesses. Still, a third option would appear to be not to choose, but to use them both. Is that indeed a viable route? To help answer this question, in the remainder of this section I discuss how the two approaches to oversight interact: how they interfere with each other and how they can best be combined.

The previous chapters have shown that the two approaches to ethical oversight can coexist in a setting without much interaction, for example when the discussions within Research Ethics Committee meetings simply alternate between the two
approaches (chapter two). So, combining the two approaches is possible in principle. However, I also found that when the two approaches to oversight lead to incompatible courses of action, i.e. when they conflict, the (dis)approval approach mostly dominates and will determine the course of action. Examples of this can be found in chapter two, where I described how a Research Ethics Committee decides to formally review and (dis)approve research proposals, even in cases where it is considered a waste of time from a quality improvement perspective. Furthermore, chapter three illustrates how a quality improvement approach to monitoring will quickly be seen as compliance monitoring (a (dis)approval approach), for instance, in case findings will be used by the Research Ethics Committee to punish researchers. Moreover, most of the scientific literature, e.g. the Oxford Textbook on Clinical Research Ethics, I encountered during my studies focuses on the (dis)approval philosophy (Emanuel et al. 2008). Legal documents and ethical guidance documents also air the spirit of (dis)approval, e.g. these documents explicitly vest Research Ethics Committees with the task of approving research.

Still, even though (dis)approval is the dominant approach, this does not mean that improving cannot, or does not, play an important role in oversight. For example, in chapter three I have described how research institutions have set up monitoring programs that work according to an improvement philosophy. Furthermore, in chapter two I described how a Research Ethics Committee employed an improvement approach that was meant to be supportive of a broader (and dominant) (dis)approval philosophy: the committee spent much time on giving researchers advice on how to improve their proposals in order to help get it approved. Using a combination of both approaches can in principle be a viable way to perform ethical oversight.

Nevertheless, whether and how to combine a (dis)approval and improvement philosophy merits careful consideration. The question of how to set up a monitoring program illustrates this aptly. In chapter three I have argued that using an improvement approach to monitor ongoing research is generally the better choice, because it will better contribute to protecting research subjects. However, I also argued that quality improvement monitoring should not be carried out by Research Ethics Committees, but by a separate office. The reason for this is that people are aware that committees are required by law to (also) work according to a (dis)approval approach. So, even if a
committee would want to monitor using a quality improvement approach, researchers will still fear that it will in practice end up in ‘policing’, i.e. compliance monitoring according to a (dis) approval philosophy. This would make it very hard for monitors to be trusted, a precondition for effective quality improvement monitoring. Following the same line of reasoning, I have argued that separate quality improvement offices should not engage in formal, i.e. (dis)approval-type, investigations of research misconduct: if they did, all their activities, including their quality improvement activities, would be perceived as policing, i.e. (dis)approval-type oversight.

So, although combining a (dis)approval and improvement philosophy within oversight bodies and within research institutions is possible and worthwhile, it is not uncomplicated. It should be carefully considered how the two approaches might interact. Explicitly distinguishing between the approaches towards researchers, for instance by having separate offices for each approach, can be helpful in this respect (Timmermans and Berg 2003).

7.8 Conclusion

In this thesis I ventured to open the black box of ethical oversight on medical research. We now return to Jan’s (the research subject mentioned in the prologue) questions about ethical oversight on research. Because we have not performed a detailed study of his case, we are not in a position to answer his first questions (whether it had been a good idea to allow an alteration of his treatment based on the MRI; whether he was informed properly; and whether in his particular case the knowledge gained by the study was worth the risks). However, we can answer his more meta-level questions: who decides about the ethical quality of research; who ensures that the ethical quality is in order; and how do they do that?

These would be my answers to him. Research Ethics Committees have the authority to decide whether to approve or disapprove research, but making approval decisions is not their only aim: they also work towards improvement of the ethical quality of research. Furthermore, ensuring that the ethical quality of research is in order is done by Research Ethics Committees, but also by others, such as governmental
bodies and offices at research institutions. Moreover, ‘ensuring’ is not the only activity in this respect: committees and other bodies also advise researchers. Currently, Research Ethics Committees and other oversight bodies focus mainly on the paperwork of research proposals and less on the actual practice of (ongoing) research, but working together with researchers to improve the practice of research is a viable alternative.

To conclude, although (dis)approving is probably closer to what people would expect a Research Ethics Committee to do and although it is indeed the dominant approach to ethical oversight, improving is a valuable alternative for supporting scientific progress and guarding the interests of research subjects. What is more, distinguishing between a (dis)approval and an improvement philosophy points to a final, more meta-level lesson, relevant to Jan and other parties interested in ethical oversight. A better understanding of the practice of ethical oversight on medical research can itself be used in two ways: to (dis)approve of current practices of oversight or to help improve those practices. My studies indicate that although the first way is likely to be dominant, the second way might be just as valuable.
References


Savulescu, J. (2002) Two deaths and two lessons: is it time to review the structure and function of research ethics committees?, *Journal of Medical Ethics*, 28, 1, 1-2.


EPILOGUE

February 1, 2008. Excerpt from an interview with a former Research Ethics Committee chair.

“The most important thing for the success of the [Research Ethics Committee is is that it] is seen as a respected agency within the institution […] If the perception of the [Research Ethics Committee] is that it is attempting to do a good thing for [the] institution, then the [people working at] the institution will help […] If [these people] see something that does not look quite proper in the conduct of research, they will call the [Research Ethics Committee] and say: 'I think there is a problem with professor so and so's research and I would like you to look into this.' If, on the other hand, [the committee is] seen […] as a police organization, then people will avoid [it]. [That is what happens when an] oversight agency develops a reputation of acting in a punitive way; not placing the highest priority on the interests of the regulated.

When I was chair of the [Research Ethics Committee], at one time I let everyone know, [that if you have] a complaint about research: bring it to me, I will attend to it immediately. [There was a] case where one of our residents called me [with a complaint about the research of one of our professors.] The two of us went together and in a courteous way I learned that what the resident had reported to me was true. I said to [the professor]: you came here from another institution and this is your first year here and perhaps this was acceptable behavior at the other institution, but it’s not here and I don't expect to ever hear of that you’ve done any such thing again. No reporting, no headlines, no nothing, very quietly.
Now, regulatory agencies despise that: how can ... you go out and, person to person, resolve the problem without creating a record? And it’s because of their insistence on documentation that the possibilities for this sort of collegial problem solving are disappearing ... I reported that case and two other cases [in a publication], with all the names changed, and said: Here is how we’re handling this at the university level and if you ever require to bureaucratize this, [to] document and report it, you will destroy this. You will dry up ... the informal monitoring system of the research environment and it will cost you millions and millions of dollars to replace this with hired monitors ... But what we have now costs us nothing."
SUPPLEMENT

Uitgangspunten inzake opslag en gebruik van (tumor) weefsel ten behoeve van aan kanker gerelateerde diagnostiek, behandeling en wetenschappelijk onderzoek

[Guidelines on storage and use of (tumour) tissue for cancer related diagnostics, treatment and scientific research.]

Corrette (M.) C. Ploem, Jean Philippe de Jong, Dick L. Willems, Sabine C. Linn, Marc J. van de Vijver, Valesca P. Retèl, Sjef (J.) K. M. Gevers, Wim. H. van Harten

I. Inleiding

Opslag/gebruik van (tumor)weefsel voor patiëntenzorg

Binnen ziekenhuizen wordt in het kader van diagnostiek bij patiënten met kanker (tumor)weefsel afgenomen (biopt) of verwijderd om dit vervolgens nader te analyseren. Dit weefsel wordt bijvoorbeeld in de vorm van een paraffineblokje of als coupe opgeslagen. Soms wordt het ook zonder bewerking ingevroren voor later gebruik. Het afgenomen (tumor)weefsel is uniek in die zin dat het niet vervangbaar is door ander (later afgenomen) weefsel.

Naast ‘traditionele’ methoden worden nieuwe diagnostische methoden ontwikkeld, zoals bijvoorbeeld ‘microarray tests’. In lijn der verwachting ligt dat uitvoering van dit soort tests ook vele jaren na de initiële behandeling nog van belang

---

1 The development of these guidelines was requested by the Netherlands Cancer Institute (NKI) in 2007. H. van Boven, M. Schmidt, F. van Leeuwen, E. Vermeulen, S. Verhoef, L. van ’t Veer, E. Vos, A. Vos, M. Pernet also contributed to their development. The guidelines formed the basis for a scientific publication: (Ploem et al. 2010).
kan zijn voor de gezondheid van de patiënt of diens eerste- en tweedegraads familieleden (bij de familieleden mogelijk ter voorkoming van ziekte).

Tegen deze achtergrond dient de huidige bewaarpraktijk binnen ziekenhuizen opnieuw te worden bezien, in die zin dat weefsel van een patiënt dat voor de uitvoering van dergelijke tests noodzakelijk is langdurig voor behandelprobleem mogelijkenheden bewaard blijft.

**Gebruik van (tumor)weefsel voor wetenschappelijk onderzoek**


Met gebruik van restweefsel voor wetenschappelijk onderzoek dienen patiënten impliciet of expliciet in te stemmen. De exacte vormgeving van hun ‘consent’ hangt samen met de wijze waarop het (tumor)weefsel (anoniem, gecodeerd, direct herleidbaar) voor onderzoek beschikbaar komt.¹

Voor wetenschappelijk onderzoek wordt niet alleen restweefsel, maar ook extra afgenomen weefsel gebruikt. Men denke daarbij bijvoorbeeld aan de situatie waarin bij een patiënt die een kankerbehandeling ondergaat (of heeft ondergaan) specifiek ten behoeve van wetenschappelijk onderzoek extra bioppen worden afgenomen. Gebruik van extra afgenomen weefsel valt (tezamen met de uitvoering van de andere onderdelen van een onderzoeksprotocol) onder de Wet medisch-wetenschappelijk onderzoek met mensen (WMO), en hiervoor geldt uit hoofde van die wet expliciete (schriftelijke) informed consent.

Voor alle duidelijkheid zij opgemerkt dat, mocht er bij de ontwikkeling van nieuwe, voor de patiënt of diens naasten van belang zijnde analysetechnieken

¹ In de onderzoekspraktijk maakt men vaak gebruik van gecodeerd weefsel. In de toepasselijke wetgeving (Wet inzake geneeskundige behandelingsovereenkomst) is echter alleen het gebruik van *anoniem* weefsel voor wetenschappelijk onderzoek geregeld (hiervoor geldt een ‘geen-bezwaarregeling’; zie art. 7:467 BW). De spelregels inzake gebruik van *gecodeerd* en *direct herleidbaar* materiaal zijn wel in zelfregulering neergelegd; zie de door de FMWV uitgebrachte Code of Conduct ‘Proper Secondary use of Human Tissue’ (Federation of Biomedical Scientific Societies 2001).
uitsluitend ‘restweefsel’ of speciaal voor wetenschappelijk onderzoek afgenomen weefsel beschikbaar zijn, daarop (door de patiënt of diens naasten) alleen een beroep kan worden gedaan indien desbetreffend weefsel in direct herleidbare vorm is opgeslagen. Weefsel dat in anonieme of gecodeerde vorm is opgeslagen is in principe – vanwege de inmiddels getroffen technische en organisatorische maatregelen bij terbeschikkingstelling van het materiaal – niet meer traceerbaar voor patiëntenzorg.

Juridisch-ethisch kader bij opslag en gebruik van lichaamsmateriaal

Dit document biedt juridisch-ethische uitgangspunten voor het reguleren van opslag en gebruik van lichaamsmateriaal, waarbij zowel de verantwoordelijkheden van instelling en hulpverlener als de rechten van patiënten en hun naasten aan bod komen. Er is bewust gekozen voor principes met een algemeen karakter opdat (andere) instellingen die met opslag en gebruik van weefsel te maken hebben ze kunnen omzetten in (lokale) richtlijnen of reglementen, toegesneden op en rekening houdend met de dagelijkse praktijk en omstandigheden binnen die instellingen. Het verdient aanbeveling dat men dergelijke instellingsrichtlijnen ter toetsing voorlegt aan een medisch-ethische commissie; die kan ook geraadpleegd worden als bij de toepassing sprake is van knelpunten of conflicten.

II. Achtergrond en globale inhoud uitgangspunten

De laatste jaren is de aandacht voor de ethische en juridische aspecten van opslag en gebruik van lichaamsmateriaal sterk gegroeid. Als gevolg daarvan zijn er steeds meer documenten gepubliceerd die aanbevelingen op dit terrein bevatten, met name waar het gaat om opslag voor researchdoeleinden. Daarbij komen de meest uiteenlopende onderwerpen aan de orde zoals beveiliging en andere privacyaspecten; de positie van minderjarigen en wilsonbekwamen; de vraag of bij opslag en gebruik sprake mag zijn van winstoogmerk en zo ja, hoe moet worden aangekeken tegen ‘benefit sharing’; de bij bewaring in acht te nemen veiligheids- en kwaliteitseisen; verstrekking van materiaal aan het buitenland; beheerstructuur van de ‘bank’ waarin het materiaal is opgeslagen; klachtrecht enz.
De hieronder weergegeven uitgangspunten beperken zich zo veel mogelijk tot de specifieke vragen die kunnen rijzen rond (tumor)weefsel. Voor de vele andere kwesties die spelen bij opslag en bewaring van lichaamsmateriaal zij verwezen naar die algemene documenten. Voor Nederland is in dit verband met name van belang de Code of Conduct ‘Proper Secondary use of Human Tissue’ (Federation of Biomedical Scientific Societies 2001).

Hoewel specifieke wetgeving met betrekking tot zeggenschap van lichaamsmateriaal vooralsnog ontbreekt (een wetsvoorstel ‘zeggenschap lichaamszorg’ is in voorbereiding), is er wel andere wetgeving die bij opslag en gebruik van lichaamszorg van betekenis is, zoals de wettelijke regeling van de geneeskundige behandelingsovereenkomst in het Burgerlijk Wetboek (WGBO), de Wet medisch-wetenschappelijk onderzoek met mensen (WMO), de Wet veiligheid en kwaliteit lichaamszorg (WVKL) en de Wet bescherming persoonsgegevens (WBP).

In internationaal verband zijn met name op het terrein van wetenschappelijk onderzoek relevante richtlijnen en aanbevelingen tot stand gekomen, waaronder: Additional Protocol to the Convention on Human Rights and Biomedicine concerning Biomedical Research (Council of Europe 2005); Recommendation on research on biological materials of human origin (Council of Europe Committee of Ministers 2006); International Declaration on Human Genetic Data (UNESCO 2003).


In genoemde juridische bronnen en ethische literatuur komt een aantal principes tot uitdrukking die aan de hierna volgende uitgangspunten ten grondslag liggen en daarin doorwerken. Het gaat in het bijzonder om de volgende vier premissen:

1. De hulpverlener dient het medisch belang van zijn patiënten te behartigen; dat is ook wat een patiënt in het algemeen van een hulpverlener mag verwachten.
In het licht van de toenemende mogelijkheden van diagnostiek op (tumor)materiaal kan bescherming van dit belang betekenen dat nadere analyse van (opgeslagen) weefsel plaatsvindt.

2. De patiënt heeft een belangrijke stem in bewaring en gebruik van zijn (tumor)weefsel voor medische doelen. Dat geldt uiteraard ook voor het afstaan van weefsel voor wetenschappelijk onderzoek.

3. Als de doelen patiëntenzorg en wetenschappelijk onderzoek conflicteren, gaat het medisch belang van de patiënt boven het belang van wetenschappelijk onderzoek. Dit geldt ook indien er bij schaarste van opgeslagen weefsel keuzen moeten worden gemaakt ten aanzien van de bestemming van het materiaal.

4. Ook het belang van naasten verdient (een zekere) bescherming, al zullen hulpverleners jegens hen doorgaans minder vergaande verplichtingen hebben dan jegens hun patiënten.1

Extra eisen stellen aan bewaring van weefsel betekent al snel extra inspanningen/kosten voor de betrokken instellingen en beroepsbeoefenaren. Bij de wijze waarop aan onderhavige uitgangspunten vorm wordt gegeven op instellingsniveau zal die extra last uiteraard een rol spelen. De onderliggende gedachte is dat men doet wat redelijkerwijs mogelij is om het gerechtvaardigd belang van de patiënt, respectievelijk diens naasten, te dienen. Zijn de daarmee gemoeide inspanningen prohibitief, dan zal in overleg met financiers ruimte moeten worden gezocht.

De indeling van de uitgangspunten is als volgt:

1. Begrippen/definities
2. Opslag/gebruik van lichaamsmateriaal voor patiëntenzorg
3. Gebruik van lichaamsmateriaal voor wetenschappelijk onderzoek

In de toelichting wordt voor zover nodig nader ingegaan op achtergrond en motivering van het betreffende uitgangspunt.

1 De hulpverlener heeft immers primair rekening te houden met (de belangen van) de patiënt met wie hij een behandelingsovereenkomst heeft of waarbij hij als hoofdverantwoordelijk voor de uitvoering van de behandeling is betrokken.
III. Uitganspunten

1. **Begrippen/definities**

1.1 Patiënt: degene bij wie het (tumor)weefsel is afgenomen.

1.2 Naasten: bloedverwanten van de patiënt tot en met de 2e graad (ouders, broers, zussen, kinderen en kleinkinderen)

1.3 Instelling: de behandel- of onderzoeksinstelling waar weefsel wordt afgenomen en/of wordt beheerd.

1.4 Behandelaar: de arts (doorgaans internist-oncoloog) met wie de patiënt de overeenkomst tot behandeling van zijn ziekte heeft gesloten dan wel degene die namens de instelling hoofdverantwoordelijk is voor zijn behandeling.

1.5 Wetenschappelijk onderzoeker: degene die binnen de instelling hoofdverantwoordelijk is voor de uitvoering van het desbetreffende onderzoek(sprotocol).

1.6 Weefsel/lichaamszament: een biopt of blokje (tumor)weefsel afkomstig van de patiënt en geschikt om gebruikt te worden voor latere diagnostiek.

1.7 Restweefsel: stuk(ge) overgebleven weefsel dat niet meer nodig is voor actuele en toekomstige patiëntenzorg en bijgevolg beschikbaar is voor andere doelen (zoals wetenschappelijk onderzoek).

1.8 Extra afgenomen weefsel: (tumor)weefsel of biopten die in het kader van een klinisch onderzoeksprotocol (uitsluitend) ten behoeve van medisch-wetenschappelijk onderzoek worden afgenomen.

1.9 Direct herleidbaar weefsel: weefsel dat zodanig (identificerend) is opgeslagen dat het rechtstreeks kan worden herleid tot de betrokkene patiënt (bijv. aan de hand van naam of initialen en geboortedatum van de betrokkene).

1.10 Indirect herleidbaar weefsel: weefsel dat zodanig (gecodeerd) is opgeslagen dat het alleen via gebruikmaking van een sleutel (bij voorkeur in beheer van een onafhankelijke derde) kan worden herleid tot de betrokkene patiënt.

1.11 Anoniem weefsel: weefsel dat zodanig (anoniem) is opgeslagen dat de mogelijkheid ontbreekt om het materiaal op een later moment weer aan de betrokkene patiënt te koppelen.
1.12 Weefselbank: verzameling van (in bijvoorbeeld paraffine of coupes opgeslagen of diepgevroren) weefsels van een groep patiënten.

1.13 Beheerder van de weefselbank: de afdeling of groep van professionals die verantwoordelijk is voor onder meer bewaring en terbeschikkingstelling van opgeslagen weefsel (bij opslag van tijdens de behandeling afgenomen weefsel: de afdeling pathologie; bij ten behoeve van wetenschappelijk onderzoek afgenomen of gebruikt weefsel: de hoofdonderzoeker of sponsor).

2. **Opslag/gebruik van lichaamsmateriaal voor patiëntenzorg**

2.1 De instelling ziet er op toe dat voldoende weefsel wordt bewaard voor (latere) diagnostiek, dat dit zorgvuldig gebeurt en dat patiënten hierover adequaat worden geïnformeerd.

Toelichting: Het in het kader van behandeling en diagnose afgenomen (tumor)weefsel dient in voldoende mate voor huidige/toekomstige diagnostiek en onderzoek bewaard te worden. Dit laatste impliceert dat a priori voldoende weefsel bij de patiënt wordt afgenomen. Indien het weefsel voorafgaand aan invoering van op deze uitgangspunten gebaseerde richtlijnen beschikbaar is gekomen en met deze bewaarpraktijk nog geen rekening kon worden gehouden, dient zo nodig (en voor zover mogelijk) alsnog (tumor)weefsel voor toekomstige diagnostiek bewaard te worden.

2.2 De instelling zorgt ervoor dat weefsel wordt bewaard zolang het (huidig of toekomstig) diagnostisch belang van de patiënt dat vordert, onverlet het recht van de patiënt om vernietiging of overdracht van zijn materiaal te verzoeken.

Toelichting: Uitgangspunt bij het vaststellen van de bewaartijd van het weefsel ten behoeve van patiëntenzorg is ‘de zorg van een goed hulpverlener’ (deze vormt in de WGBO ook het uitgangspunt voor het bewaren van medische gegevens). Het is van belang dat die bewaartermijn zodanig ruim is dat behalve aan het belang van de patiënt zelf na diens overlijden ook tegemoet kan worden gekomen aan de belangen van zijn naasten (men denke aan (vroelege) opsporing van famillaire/erfelijk tumor of
ziekten). De patiënt behoudt – naar analogie van diens vernietigingrecht met betrekking tot zijn medische gegevens (zie art. 7: 455, eerste lid BW) – gedurende die bewaarperiode het recht het (primair in zijn belang bewaarde) weefsel te laten vernietigen (zie verder uitgangspunt 2.9). Indien zijn gezondheidsbelang dat vordert, kan een patiënt verzoeken zijn weefsel over te dragen naar een andere instelling (zie ook 2.10). Ten slotte kan een patiënt zijn materiaal ter beschikking stellen (doneren) voor wetenschappelijk onderzoek. Zie voor de positie van nabestaanden uitgangspunt 2.12.

2.3 Na het verstrijken van de bewaartermijn wordt weefsel dat nog over is vernietigd, tenzij het diagnostisch belang van naasten dan wel het algemeen belang van de wetenschap zich daartegen verzet. Toelichting: In principe volgt na het aflopen van de bewaartermijn vernietiging van het weefsel. Niettemin zal enige tijd voor het aflopen van de bewaartermijn moeten worden bezien in hoeverre belangen van naasten dan wel de wetenschap zich tegen vernietiging verzetten. In dat laatste geval dient het weefsel wederom op zorgvuldige wijze voor een nader te bepalen doel en termijn te worden bewaard. Een voorgenomen besluit van de instelling om (bepaald) weefsel langer te bewaren in het belang van wetenschappelijk onderzoek dient ter toetsing aan de binnen de instelling functionerende medisch-ethische toetsingscommissie (METC) te worden voorgelegd. De METC gaat onder meer na of de doelstelling van bewaring helder is, de wijze waarop het weefsel wordt bewaard verantwoord is (kwaliteitsseisen, privacybescherming e.d.) en de rechten van de betrokken patiënten zich daartegen niet verzetten.

2.4 Na beëindiging van de behandelrelatie draagt de instelling waar het weefsel wordt bewaard ervoor zorg dat het materiaal desgevraagd (en met instemming

1 De Gezondheidsraad heeft gepleit voor opslag van gegevens voor tenminste een periode van 30 jaar (Gezondheidsraad 2004), maar een langere bewaartermijn van gegevens (en lichaamsmateriaal), bijv. 100 jaar of langer, lijkt – gelet op de huidige stand van behandeling en (preventief) onderzoek – van belang voor (in ieder geval) patiënten met een ziekte waarvoor mogelijk een erfelijke aanleg bestaat (zoals kanker).
van de patiënt) ter beschikking wordt gesteld aan nieuwe, mogelijk in een andere instelling werkzame behandelaars van de patiënt (of van diens naasten).

Toelichting: Het is goed voorstelbaar dat op het moment dat een patiënt (of diens naasten) belang heeft (hebben) bij verdere analyse van afgenomen (tumor)weefsel (bijv. voor het kiezen van een adequate preventieve strategie, ter voorkoming van een recidief/metastasen) de behandelingsovereenkomst met de oorspronkelijke instelling/behandelaar is beëindigd. In die situatie zouden nieuwe behandelaars of andere, bij de zorg voor de patiënt (of diens naasten) betrokken hulpverleners over (een deel van) het bewaarde weefsel moeten kunnen beschikken voor zover dat voor diagnostische doeleinden noodzakelijk is en de patiënt hierom verzoekt respectievelijk hiermee instemt. De kring van gebruikers van opgeslagen materiaal is dus potentieel ruim en strekt zich uit over de grenzen van de instelling.

2.5 De instelling stelt een protocol en/of reglement op waarin alle relevante aspecten rond opslag en bewaring van weefsel worden geregeld.

Toelichting: Ter ondersteuning van een goede en transparante bewaarpraktijk stelt de instelling een ‘bewaarprotocol’ op waarin alle relevante aspecten rond het bewaren van het weefsel worden geregeld, waaronder de wijze van opslag (zoals bij voorkeur fysiek gescheiden opslag van voor diagnostiek bestemd weefsel en ‘restweefsel’), de bewaartijd, de te volgen procedure wanneer er vanuit het wetenschappelijk onderzoek een beroep op het weefsel wordt gedaan, de maatregelen die nodig zijn om een verantwoorde bewaring te waarborgen, de wijze waarop met verzoeken van patiënten (of hun naasten) met betrekking tot hun weefsel (zie met name uitgangspunten 2.7 tot en met 2.12) wordt omgegaan, etc.

2.6 De behandelaar houdt de patiënt – voor zover dit voortvloeit uit ‘de zorg van een goed hulpverlener’ – op de hoogte van nieuwe mogelijkheden om diens weefsel te onderzoeken voor zover deze voor de patiënt van klinisch belang kunnen worden geacht, tenzij dit in redelijkheid niet mogelijk is of indien de patiënt heeft aangegeven daarover niet te willen worden geïnformeerd.

Toelichting: De behandelaar zorgt ervoor dat de bij hem in behandeling zijnde patiënten van wie weefsel conform deze richtlijn beschikbaar is, op de hoogte worden

Overigens kan zich de situatie voordoen waarin het informeren van de patiënt in redelijkheid niet mogelijk is (de desbetreffende patiënt is verhuisd of het is onkies de betrokken patiënt vanwege diens gezondheidstoestand opnieuw met zijn ziekte te confronteren) of waarin de patiënt zelf heeft aangegeven van zulke informatie geen kennis te willen nemen.

2.7 Voor zover een nieuwe test deel uitmaakt van de professionele standaard, behoort deze de te worden uitgevoerd als de patiënt daarom verzoekt.

Toelichting: De behandelaar komt tegemoet aan een verzoek van een patiënt een bepaalde test bij hem uit te voeren indien deze tot de geldende medische standaard van zorg kan worden gerekend (dit impliceert dat deze test het experimentele traject heeft doorlopen). Is dat het geval, dan zal de uitvoering ervan overigens naar men mag aannemen als geaccepteerde medisch-specialistische zorg verzekerd zijn uit hoofde van de Zorgverzekeringswet.

Ook als een test nog niet tot de geldende medische standaard kan worden gerekend, kunnen behandelaars van mening zijn dat uitvoering aangewezen is. Mocht over het nut van de uitvoering van een diagnostische test tussen patiënt en behandelaar verschil van inzicht bestaan en weigert de behandelaar de test uit te voeren, dan kan de patiënt verzoeken het tumormateriaal over te dragen aan een andere instelling waar men wel bereid is de test uit te voeren (zie 2.10).

2.8 Over de opslag en langdurige bewaring van weefsel voor toekomstige therapeutische doeleinden, over zijn rechten, en over de positie van zijn naasten wordt de patiënt expliciet geïnformeerd.

Toelichting: Cruciaal voor een zorgvuldige omgang met binnen de instelling behandelde patiënten is adequate informatieverstrekking over de in deze uitgangspunten beschreven bewaringspraktijk (zie ook het in uitgangspunt 2.5
genoemde bewaarprotocol of reglement), zijn daaraan gekoppelde rechten, en de positie van zijn directe naasten. Van belang is dat betrokkenen niet alleen van hun rechten zoals in deze paragraaf omschreven op de hoogte worden gesteld, maar ook dat zij in algemene zin worden geïnformeerd over opslag en gebruik van ‘restmateriaal’ voor wetenschappelijk onderzoek.
Zoals aangegeven in uitgangspunt 2.1 is het primair de instelling die voor een dergelijke informatieverstrekking verantwoordelijk is (men denke aan een aparte folder/internetpagina over opslag en gebruik van tumorweefsel).

2.9 De patiënt kan verzoeken zijn weefsel te vernietigen1 dan wel aan de wetenschap ter beschikking te stellen.
Toelichting: Langdurige bewaring van onbewerkt weefsel dient primair het gezondheidsbelang van betrokken patiënten (en zijn naasten). Het is evenwel niet ondenkbaar dat een patiënt aan (langdurige) bewaring van zijn materiaal geen behoefte (meer) heeft. Hij kan dan ofwel om vernietiging van het weefsel verzoeken, ofwel zijn weefsel ter beschikking stellen (doneren) aan de wetenschap (zie ook uitgangspunt 2.2 en de toelichting daarop).

2.10 Op verzoek van de patiënt wordt het weefsel overgedragen aan een andere instelling.
Toelichting: Een verzoek tot overdracht van weefsel kan aan de orde zijn indien de patiënt inmiddels elders onder behandeling is, ten behoeve van een naaste die elders onder behandeling is, of wanneer de behandelaar aangeeft de test waarom de patiënt vraagt niet te willen of kunnen uitvoeren (zie ook 2.7).

1 Het in art. 7: 455, lid 1 BW neergelegde ‘vernietigingsrecht’ van patiënten is overigens niet absoluut: op een dergelijk verzoek hoeft niet te worden ingegaan indien redelijkerwijs aannemelijk is dat verdere bewaring van medische gegevens van aanmerkelijk belang is voor een ander dan de patiënt, zoals een naaste van de patiënt die (mogelijk) een erfelijke ziekte heeft. In de rede ligt dat bij een verzoek tot vernietiging van weefsel door een patiënt door de hulpverlener eveneens wordt bezien in hoeverre een aanmerkelijk belang van nabestaanden zich daartegen verzet.
2.11 Naasten kunnen ter behartiging van hun eigen diagnostisch belang vragen om overdracht van (een deel van) het materiaal mits de patiënt daarmee instemt.
Toelichting: Opslag (na overdracht) van tumorweefsel binnen een andere instelling en/of diagnostisch gebruik van weefsel van de patiënt door diens naasten vereist expliciete (schriftelijke of mondelinge) toestemming van die patiënt.

2.12 Na overlijden van de patiënt kunnen naasten de instelling waar het materiaal wordt bewaard verzoeken om overdracht van (een deel van) het overgebleven tumormateriaal van de overledene aan hun behandelaar voor zover dat in het kader van hun eigen behandeling nodig is.
Toelichting: Nabestaanden kunnen, voor zover nodig voor diagnostiek bij henzelf, vragen om de beschikbaarstelling van het materiaal aan hun behandelaar.

3. **Gebruik van lichaamsmateriaal voor wetenschappelijk onderzoek**
3.1 De instelling ziet erop toe dat voor de uitvoering van wetenschappelijk onderzoek uitsluitend ‘restweefsel’ of speciaal ten behoeve van wetenschappelijk onderzoek afgenomen *extra* weefsel wordt gebruikt.
Toelichting: Conform uitgangspunt 2.1 mag gebruik van (tumor)weefsel voor wetenschappelijk onderzoek niet ten koste gaan van gebruik in het belang van de patiënt zelf; hieruit volgt dat voor de uitvoering van een studie alleen restweefsel (zie 1.7) of *extra* afgenomen weefsel (zie 1.8) mag worden gebruikt.
Mocht er vanwege schaarste van het afgenomen (tumor)materiaal geen ‘restweefsel’ beschikbaar zijn, dan mag de wetenschap – tenzij de patiënt overleden is en er geen naaste familieleden zijn voor wie bewaring van belang kan zijn – op het voor patiëntenzorg gereserveerde (en bewaarde) weefsel geen beroep doen. Wanneer binnen de instelling geen gescheiden opslag plaatsvindt van voor patiëntenzorg bestemd weefsel en voor andere doelen beschikbaar ‘restweefsel’ (dit is overigens wel aan te bevelen) dient op de een of andere wijze te zijn gewaarborgd dat steeds voldoende tumorweefsel van een patiënt voor diens behandeling bewaard blijft.

3.2 Wanneer een onderzoeker direct herleidbaar of gecodeerd tumorweefsel analyseert, en er sprake is van een individuele bevinding die van belang is voor
de betrokken patiënt en/of diens naaste familieleden (kinderen, zussen/broers, ouders), wordt de patiënt daarvan via de behandelend arts of daarvoor in de plaats tredend hulpverlener op de hoogte gesteld, tenzij de patiënt eerder te kennen heeft gegeven van zulke (incidentele) bevindingen niet op de hoogte te willen worden gesteld.

Toelichting: Wetenschappelijk onderzoek kan soms leiden tot individuele bevindingen die voor de betrokken onderzoekssubjecten van direct klinisch belang zijn. De consensus in de (internationale) ethische literatuur is dat een patiënt geïnformeerd moet worden over zulke incidentele bevindingen, mits deze voldoende betrouwbaar en valide zijn. Uit die literatuur komt voorts naar voren dat patiënten bij klinisch relevante bevindingen of anderszins belastende informatie er de voorkeur aan geven daarvan door hun behandelaar (met wie zij een vertrouwensrelatie hebben) op de hoogte te worden gesteld.

3.3 De patiënt of proefpersoon wiens (tumor)weefsel in herleibare of gecodeerde vorm in het kader van wetenschappelijk onderzoek wordt geanalyseerd, dient – wanneer hij informatie krijgt over zaken als onderzoeksdoel, bewaartijd van het weefsel, beveiliging en privacybescherming – ook te worden geïnformeerd over de mogelijkheid dat uit de analyse van zijn weefsel bevindingen naar voren kunnen komen die voor hem (en/of zijn naasten) van belang zijn. Hiermee samenhangend dient de patiënt of proefpersoon op de hoogte te zijn (of te worden gesteld) van zijn recht aan te geven van die (of bepaalde) bevindingen niet op de hoogte te willen worden gebracht ('recht op niet-weten').

Toelichting: Van belang is dat patiënten/proefpersonen zich realiseren dat uit analyse van hun weefsel in het kader van wetenschappelijk onderzoek incidenteel ook bevindingen naar voren kunnen komen die naar het oordeel van de onderzoeker voor de patiënt/proefpersoon en/of zijn naasten van direct belang kunnen zijn en, mits voldoende betrouwbaar en klinisch relevant, via de behandelaar aan hen worden meegedeeld, tenzij ze hebben aangegeven van die bevindingen niet op de hoogte te willen worden gesteld.
References


Council of Europe Committee of Ministers (2006) Recommendation of the Committee of Ministers to member states on research on biological materials of human origin.


SUMMARY

Chapter one: Introduction

This thesis studies how independent ethical oversight on medical research with human subjects functions in practice. In the introductory chapter, I explain why there is a need to study ethical oversight and what my approach is for doing so. In most countries, a system for ethical oversight on medical research, with Research Ethics Committees as its cornerstone, has been in place for many years. Such a system aims to ensure that the ethical quality of research is in order and that subjects are protected. The ethical quality of a study depends on whether the study is in line with relevant moral principles: the social value of research, the scientific validity, a fair subject selection, a favorable ratio of risks and burdens to subjects and benefits for society, independent ethical review, informed consent of subjects, and respect for enrolled subjects. I define ethical oversight as the activities of independent organizations that work on the ethical quality of medical research.

Although Research Ethics Committees are generally considered to play an important role in the system of ethical oversight, they have also been criticized. There has been frustration about the efficiency of committees and some feel that they unnecessarily impede the progress of science. At the same time, there have also been doubts about whether Research Ethics Committees provide adequate protection for research subjects. In reality however, little is known about the functioning of Research Ethics Committees or the broader practice of ethical oversight: the oversight system is a ‘black box’. The central aim of this thesis is to open this black box and to help understand how ethical oversight, especially by Research Ethics Committees, works in
practice. A secondary aim of this thesis is to help improve the functioning of ethical oversight.

To achieve these aims, this thesis addresses the following five research questions: How do Research Ethics Committees evaluate proposals for medical research during their meetings? Do Research Ethics Committees and research institutions oversee whether the actual conduct of research is ethical, and if so, how? Do approved studies actually lead to scientific knowledge and how can Research Ethics Committees oversee that they do? How to handle cases where patients’ medical interests conflict with the interest of scientific progress? And how can oversight bodies make sure that the benefits of research are distributed fairly among the population, thus serving the public good?

To enter the black box of ethical oversight and answer these questions, I have looked for the contrasts – oppositions that emphasize differences – within the practice of ethical oversight, in order to describe different ‘ways of doing good’. In each of the following five chapters I take up one of the research questions and analyze the contrasts within them. In the concluding chapter, chapter seven, I take my analysis of ethical oversight one step further and synthesize the contrasts I uncovered in the preceding chapters into a more general understanding of ethical oversight.

Chapter two: Ethical review from the inside

In chapter two I delve into one of the most important, but at the same time least transparent elements of the system of ethical oversight: the deliberations that take place during Research Ethics Committee meetings. I ask how Research Ethics Committees evaluate the proposals for medical research in practice, during their meetings. Finding an answer to this question could help improve how Research Ethics Committees protect the interests of both subjects and science. By sitting in on committee meetings and analyzing the discussions I discovered that committees are involved in two repertoires of evaluation: a repertoire that focuses on rules and judgments, and a repertoire that focuses on knowledge production and advice. I suggest that although the former repertoire is closer to what many expect from Research Ethics Committees, using the
two repertoires in conjunction is worth the while, because it helps researchers improve the ethical quality of research proposals.

**Chapter three: Compliance or quality improvement?**

In chapter three I move beyond the review of research proposals, and study what ethical oversight on the actual conduct of research amounts to in practice. I ask how Research Ethics Committees and research institutions monitor the conduct of research and why they have arranged it this way. Getting insight in monitoring practices is important because monitoring ongoing research has been proposed as an additional way of improving the protection of the rights and welfare of research subjects. I studied Research Ethics Committees and research institutions in the U.S. because independent monitoring programs have been in place there for several years, unlike The Netherlands. My analysis showed that monitoring programs varied considerably, but gravitated towards two general types: *compliance monitoring*, which focuses on documentation, and can amount to disciplining researchers and requiring mandatory corrective actions; and *quality improvement monitoring*, which focuses more on actual research conduct, and can result in feedback to both researchers and the research institution on how to improve the research process. I argue that quality improvement monitoring is the better choice because it helps foster trust between researchers and Research Ethics Committees, leading to a better protection of the interests of research subjects.

**Chapter four: Publication rate of clinical studies**

In chapter four I further narrow down the question of how ethical oversight can influence the conduct of research. I investigate to what extent research studies lead to scientific progress, i.e. publications, and how Research Ethics Committees could ensure that they do. I ask to what extent study results are published, and whether a committee could predict failure to publish already during ethical review. Failure to publish is a
grave way of treading research subjects’ interests, since failure to publish makes research subjects’ efforts go in vain and can bias the scientific literature. Prediction of a failure to publish could thus give Research Ethics Committees an important tool with which to better do justice to research subjects. I found that almost half of the studies that actually included research subjects remain unpublished. Furthermore, by comparing studies that have been published to those that were not, I found that studies that had a problematic review process and studies that aimed to benefit patients directly (as opposed to fundamental research), was associated with publication failure. Research ethics committees could use this information to monitor whether studies lead to publication, and for discussing their worries with researchers during the review process in order to prevent non-publication.

Chapter five: Tumour tissue – who is in control?

In chapter five I study oversight on the use and storage of human tissue and describe the development of a guideline for the management of patients’ residual (tumor) tissue. Residual tissue is often stored for research purposes, but can also sometimes serve clinical ends at a later moment. This can lead to a conflict of interests between patients’ and research interests. This came to the fore when a woman, previously treated for breast cancer, requested her physician to have a new genetic test performed on her residual tumor tissue which had been stored for research. As guidance was lacking for how to handle this case and how to weigh the interests of a patient against those of research, developing a guideline seemed appropriate. The analysis conducted in this study showed that such a guideline should take four ethical principles into account: the responsibility of health care providers to provide good clinical care; the rights of patients regarding their bodily material, removed or not; the relative rights of family members regarding this material; and the overriding interest of patients’ medical interests over the interests of scientific research in cases where they conflict. The practical implications for the management of human tissue were also explored, including the practicalities of storing sufficient tissue for future clinical usage.
Chapter six: Justice in clinical research methods

In chapter six I examine how considerations of justice can be relevant to ethical oversight and what implications this holds for research. I explore the idea that the choice for a specific research methodology can affect whether the benefits of research results are distributed fairly. Using the case of hypertension management as an example, I argue that three elements of the current ‘gold-standard’ of research methods – the randomized controlled trial (RCT) can potentially lead to an unfair distribution: (1) RCTs tend to standardize study groups, and results in standardized study groups can be difficult to extrapolate to disadvantaged groups outside a study; (2) RCTs standardize the delivery of health care within studies, and disadvantaged patient groups outside a study may have more difficulties adapting to these standardized interventions; and (3) the focus of RCTs on standardized interventions directs medical research away from more complex interventions such as lifestyle interventions, which are needed especially by disadvantaged patients. I conclude that, although standardization can help to generate valid knowledge, in particular cases it can also mean that less relevant medical knowledge will become available for disadvantaged patient groups. I further argue that less standardization in RCTs could be beneficial, and that research methods that do not have a strong tendency to standardize, such as qualitative methods, are needed to generate relevant knowledge for disadvantaged groups. These are considerations that could be taken into account in ethical oversight.

Chapter seven: Discussion

In the concluding chapter, I work towards a deeper understanding of ethical oversight, based on a synthesis of the contrasts uncovered in the studies in this thesis. I argue that there are two philosophies for doing good in ethical oversight. First, ‘(dis)approving’, which aims to ensure that research is ethical and works towards (dis)approval, is carried out within relationships based on authority, and focuses on documentation. And second, ‘improving’, which aims to improve the ethical quality of research by giving advice, is carried out within relationships based on equality, and focuses on practice.
After reflecting on the research methods used in this thesis I discuss the strengths and weaknesses of both approaches. Subsequently, I discuss how the two approaches to oversight interact, and how they can best be combined. In conclusion, I articulate the main lessons that can be drawn from this thesis. Although (dis)approving is closer to what people expect from Research Ethics Committees and is the dominant approach to ethical oversight, improving is a valuable alternative for supporting scientific progress and guarding the interests of research subjects.
Hoofdstuk één: Inleiding

In dit proefschrift beschrijf ik mijn onderzoek naar de praktijk van ethisch toezicht op medisch wetenschappelijk onderzoek met proefpersonen. In het inleidende hoofdstuk leg ik uit waarom het wenselijk is te onderzoeken hoe ethisch toezicht in de praktijk functioneert en wat mijn aanpak daarvoor is. De meeste landen hebben al vele jaren een systeem voor toezicht op de ethische aspecten van medisch onderzoek, met Medisch-Ethische Toetsingscommissies als hoekstenen van dat systeem. Deze toezichtsysteem dienen ervoor te zorgen dat onderzoek van voldoende ethische kwaliteit is en dat proefpersonen daarmee beschermd worden tegen de gevaren van deelname aan onderzoek. Of onderzoek ethisch is hangt af van een aantal morele principes: het maatschappelijk belang van onderzoek, de wetenschappelijke validiteit, een rechtvaardige selectie van proefpersonen, een goede verhouding tussen enerzijds de risico's en belasting voor proefpersonen en anderzijds het maatschappelijk belang, een geïnformeerde toestemming van proefpersonen en tot slot, een respectvolle omgang met proefpersonen. Onder ‘ethisch toezicht’ versta ik in dit proefschrift: de activiteiten van onafhankelijke organisaties gericht op de ethische kwaliteit van medisch wetenschappelijk onderzoek met proefpersonen.

Ondanks dat Medisch-Ethische Toetsingscommissies over het algemeen gezien worden als belangrijke spelers in het toezichtsysteem, worden ze ook bekritiseerd. Er wordt getwijfeld aan de doelmatigheid van Toetsingscommissies en er wordt gedacht dat ze de vooruitgang van de wetenschap onnodig belemmeren. Tegelijkertijd zijn er ook zorgen over de bescherming van proefpersonen door Toetsingscommissies. Deze
kritieken worden echter vaak niet onderbouwd met feiten. Er is weinig betrouwbare informatie over het daadwerkelijke functioneren van Toetsingscommissies of de bredere praktijk van ethisch toezicht; het systeem is een ‘black box’. Het hoofddoel van het onderzoek in dit proefschrift is deze ‘black box’ open te maken en inzicht te geven in hoe ethisch toezicht, in het bijzonder door Medisch-Ethische Toetsingscommissies, in de praktijk werkt. Een secundaire doelstelling is te helpen het toezichtsysteem te verbeteren.

Om deze doelstellingen te verwezenlijken gaat dit proefschrift in op de volgende vijf onderzoeksvragen: Hoe evalueren Medisch-Ethische Toetsingscommissies voorstellen voor medisch onderzoek tijdens hun vergaderingen? Houden Toetsingscommissies en onderzoeksinstellingen ook toezicht op de ethische kwaliteit van de uitvoering van onderzoek, en zo ja, hoe? Leiden onderzoeken daadwerkelijk tot wetenschappelijke kennis en hoe kan een Toetsingscommissie daar toezicht op houden? Wat dient te worden gedaan in gevallen waar de medische belangen van patiënten botsen met het belang van wetenschappelijke vooruitgang? En hoe kunnen toezichthoudende organisaties zorgen voor een rechtvaardige verdeling van de baten van wetenschappelijke kennis?

Om de ‘black box’ van ethisch toezicht open te maken en de onderzoeksvragen te kunnen beantwoorden heb ik binnen de praktijk van ethisch toezicht gezocht naar contrasten: scherpe tegenstellingen die inzicht geven in diverse manieren van het goede doen. In de volgende vijf hoofdstukken behandelt ik telkens één van de onderzoeksvragen en onderzoek de desbetreffende contrasten. In het afsluitende hoofdstuk, hoofdstuk zeven, maak ik een overkoepelende analyse van de deelonderzoeken uit de eerdere hoofdstukken en distileer ik uit de gevonden contrasten een dieper inzicht in wat het is om het goede te doen in ethisch toezicht op onderzoek.

**Hoofdstuk twee: Achter de schermen bij ethische toetsing**

In hoofdstuk twee ga ik dieper in op een van de belangrijkste, maar tegelijkertijd ook een van de minst transparante onderdelen van het toezichtsysteem voor medisch
onderzoek: de discussies die plaatsvinden tijdens de vergaderingen van Medisch-Ethische Toetsingscommissies. Ik onderzoek hoe een Toetsingscommissie in de praktijk, gedurende haar vergaderingen, voorstellen voor medisch onderzoek evalueert. Kennis hierover zou Toetsingscommissies kunnen helpen bij het behartigen van de belangen van proefpersonen en de wetenschap. Door commissievergaderingen bij te wonen en de discussies die daar plaatsvonden te analyseren ontdekte ik dat de Toetsingscommissie twee handelingsrepertoires gebruikte bij het evalueren van onderzoeksvoorstellen: enerzijds een repertoire dat zich richt op regels en oordelen en anderzijds een repertoire dat zich richt op kennisproductie en advies. Ondanks dat het eerste repertoire dichter ligt bij wat mensen verwachten dat Medisch-Ethische Toetsingscommissies doen, is het combineren van deze twee repertoires de moeite waard omdat het onderzoekers kan helpen de ethische kwaliteit van hun onderzoek te verbeteren.

Hoofdstuk drie: Monitoren – naleving of kwaliteitsverbetering?

In hoofdstuk drie richt ik me op de fase na de toetsing van onderzoeksvoorstellen en kijk ik naar de daadwerkelijke uitvoering van medisch onderzoek. Ik onderzoek hoe Medisch-Ethische Toetsingscommissies en onderzoeksinstellingen in de praktijk toezicht houden op de uitvoering van onderzoek ('monitoren'). Inzicht in deze praktijken is van belang, omdat het monitoren van onderzoek een belangrijke bijdrage kan leveren aan de bescherming van de rechten en het welzijn van proefpersonen. Ik heb voor dit onderzoek gekeken naar Medisch-Ethische Toetsingscommissies en onderzoeksinstellingen in de Verenigde Staten, omdat programma’s voor het monitoren van onderzoek daar al enige jaren in zwang zijn – dit in tegenstelling tot Nederland. Mijn analyse laat zien dat er aanzienlijke verschillen bestaan tussen programma’s, maar ook dat ze ingedeeld kunnen worden in twee types. Het eerste type noem ik ‘nalevings-monitoren’. Nalevings-monitoren richt op de onderzoeksdocumentatie en kan uitmonden in het eisen van herstelacties of het straffen van onderzoekers. Het tweede type noem ik ‘kwaliteitsverbeterings-monitoren’. Kwaliteitsverbeterings-monitoren richt zich meer op de daadwerkelijke uitvoering van onderzoek en kan resulteren in lessen voor onderzoekers en de onderzoeksinstelling over hoe het onderzoeksproces
verbeterd kan worden. Ik betoog dat kwaliteitsverbeterings-monitoren de voorkeur verdient, omdat deze vorm van monitoren kan helpen een sfeer van vertrouwen tussen onderzoekers en Medisch-Ethische Toetsingscommissies te cultiveren, wat kan leiden tot een betere bescherming van proefpersonen.

**Hoofdstuk vier: Hoe vaak worden klinische studies gepubliceerd?**

In hoofdstuk vier kijk ik in meer detail naar toezicht op de uitvoerende fase van medisch onderzoek. Ik stel de vraag in hoeverre onderzoek tot wetenschappelijke vooruitgang leidt en hoe daar toezicht op gehouden kan worden. Ik heb daarom onderzocht hoe vaak onderzoeksresultaten tot publicaties leiden in wetenschappelijke tijdschriften, en of een Medisch-Ethische Toetsingscommissie dit tijdens het toetsen van een onderzoeksvoorstel al zou kunnen voorspellen. Verzuimen onderzoeksresultaten te publiceren doet immers geen recht aan de inspanningen van proefpersonen. Bovendien kan het niet-publiceren van resultaten de wetenschappelijke literatuur vertekenen. Het kunnen voorspellen van al of niet publiceren zou Toetsingscommissies een extra instrument in handen kunnen geven om de belangen van proefpersonen te beschermen. Ik ontdekte dat bijna de helft van de onderzoeken niet gepubliceerd werden. Bovendien bleek dat onderzoeken waarvan het toetsingsproces moeizaam was geweest en onderzoeken die direct ten goede hadden kunnen komen aan proefpersonen (toepast onderzoek) relatief veel vaker niet gepubliceerd werden. Toetsingscommissies zouden dit soort voorspellende informatie kunnen gebruiken om er meer gericht op toe te zien of onderzoek gepubliceerd wordt. Toetsingscommissies zouden ook tijdens het toetsingsproces hun zorgen over niet-publiceren kunnen bespreken met de onderzoeker in kwestie, om daarmee niet-publiceren te voorkomen.

**Hoofdstuk vijf: Tumorweefsel – wie gaat erover?**

In hoofdstuk vijf richt ik me op het toezicht op het gebruik en de bewaring van humaan weefsel en beschrijf ik de ontwikkeling van een richtlijn voor het beheer van rest-
SAMENVATTING

(tumor)weefsel van patiënten. Rest-weefsel wordt vaak opgeslagen voor onderzoeksdoeleinden, maar kan in sommige gevallen op een later tijdstip ook nog van klinische betekenis zijn. Hierbij kunnen de belangen van patiënten tegenover die van onderzoekers komen te staan. Een dergelijke situatie ontstond toen een vrouw, die eerder behandeld was voor borstkanker, aan haar arts vroeg een nieuwe genetische test uit te voeren op haar rest-tumorweefsel dat opgeslagen lag voor onderzoek. Omdat er voor deze situatie geen direct toepasbare richtlijnen beschikbaar waren die konden helpen bij het afwegen van de belangen van een patiënt tegenover onderzoek, leek het nuttig hiervoor een specifieke richtlijn te ontwikkelen. Een analyse van relevante literatuur en bestaande richtlijnen gaf aan dat er in toekomstige, vergelijkbare gevallen rekening gehouden zou moeten worden met vier ethische principes: de verantwoordelijkheid van zorgverleners om goede zorg te leveren; de rechten van patiënten ten aanzien van hun lichaamsmateriaal, verwijderd of niet; de afgeleide rechten van familieleden ten aanzien van het materiaal; en het doelstellinggevende belang van de gezondheid van een patiënt ten opzichte van wetenschappelijke belangen. De praktische implicaties van het beheer van weefsel worden in dit hoofdstuk nader geëxplooreerd, onder meer de praktische aspecten van het opslaan van voldoende weefsel voor toekomstig klinisch gebruik.

Hoofdstuk zes: Rechtvaardige methodes voor klinisch onderzoek

In hoofdstuk zes analyseer ik hoe het principe van verdelende rechtvaardigheid relevant kan zijn voor medisch onderzoek. Ik onderzoek hoe de keuze voor een bepaalde onderzoeksmethodologie invloed kan hebben op een rechtvaardige verdeling van de baten van de gegenererde wetenschappelijke kennis. Ik betoog dat drie onderdelen van de huidige ‘gouden standaard’ voor medische onderzoeksmethodes – het gerandomiseerde, gecontroleerde onderzoek (RCT) – kunnen leiden tot een onrechtvaardige verdeling: (1) RCTs standaardiseren onderzoeksgroepen, wat de resultaten soms moeilijk te vertalen maakt naar achtergestelde groepen buiten het onderzoek; (2) RCTs standaardiseren de zorgverlening binnen onderzoek, en achtergestelde groepen buiten het onderzoek kunnen meer moeite hebben om zich aan
deze gestandaardiseerde zorg aan te passen; en (3) de nadruk die RCTs leggen op
gestandaardiseerde interventies leidt ertoe dat er binnen onderzoek minder aandacht is
voor complexere interventies zoals leeftijdsinterventies, precies het soort interventies
dat patiënten uit achtergestelde groepen goed kunnen gebruiken. Hoewel
standaardiseren kan helpen om valide kennis te vergaren, kan het er dus ook toe leiden
dat er voor achtergestelde patiëntengroepen minder relevante kennis beschikbaar komt.
Het zou kunnen helpen om minder te standaardiseren in RCTs, en bovendien zouden
onderzoeksmethodes met een minder sterke hang naar standaardisering, zoals
kwalitatieve methodes, kunnen helpen specifieke kennis voor achtergestelde groepen te
vergaren. Dit zijn overwegingen die een plek dienen te krijgen bij ethisch toezicht op
medisch onderzoek.

**Hoofdstuk zeven: Discussie**

In het afsluitende hoofdstuk smeed ik de resultaten van de deelonderzoeken in dit
proefschrift samen tot een dieper inzicht in ethisch toezicht op medisch onderzoek. Ik
betoog dat er twee filosofieën zijn voor hoe je het goede kan doen in ethisch toezicht.
Ten eerste ‘goed- of afkeuren’: dit heeft als doel de ethische kwaliteit van onderzoek te
borgen en werkt toe naar goed- of afkeuring, kent werkrelaties gebaseerd op autoriteit,
en richt zich op documentatie. Ten tweede ‘verbeteren’: dit heeft als doel de ethische
kwaliteit van onderzoek te verbeteren door adviezen te geven, kent werkrelaties
gebaseerd op gelijkwaardigheid, en richt zich op de praktijk. Na een reflectie op de
onderzoeksmethodes die gebruikt zijn voor dit proefschrift, bespreek ik de sterke en
zwakke kanten van beide filosofieën voor ethisch toezicht. Vervolgens bespreek ik de
wisselwerking tussen beide aanpakken, en hoe ze het beste gecombineerd kunnen
worden. Ik sluit af met het trekken van de belangrijkste conclusies uit dit proefschrift.
Ondanks dat goed- of afkeuren dichter ligt bij wat mensen van Medisch Ethische
Toetsingscommissies verwachten en dat het de dominante aanpak is voor ethisch
toezicht, is verbeteren een waardevol alternatief om wetenschappelijke vooruitgang te
ondersteunen en de belangen van proefpersonen te behartigen.
ACKNOWLEDGEMENTS

In writing my PhD thesis, I received help from many others. Here, I would like to thank all of those who have kept me on track and also those who provided valuable diversions. In particular I want to express my gratitude to the following persons:

Dick, both my promotor and daily supervisor, thank you for allowing me to follow my own path, while providing guidance at crucial moments. I especially appreciate your help in arranging my visiting scholarship to the U.S. and your lenience in allowing me to study medicine ‘on the side’. In my letter of application to this PhD project I said I wanted to combine empirical and philosophical methods. I got what I bargained for: your emphasis on pluralism in doing ethics has stimulated me to expand my methodological horizon.

Myra, although you were involved neither at the end, nor at the beginning of my project, you supported me during the most important part: the middle phase. I have learned a great deal from you on qualitative analysis, structuring manuscripts and managing the intricacies of working as a PhD/scientist.

Gerben, working with you has made me appreciate both the complexity and value-based character of statistical analysis. You have been a great help in building my first multi-variate model.

Corrette and Sjef, thank you for a fruitful collaboration on developing a guideline on the storage of tumour tissue and for providing insight into legal thinking.

I thank the board of the Academic Medical Center for financially supporting my PhD project and for its appreciation of the merits of studying the practice of ethical oversight; The Netherlands Cancer Institute for sponsoring a study to develop a
guideline on the storage of tumour tissue; and the Van Walree Fund for sponsoring a visiting scholarship to the U.S.

The department of General Practice and its people I thank for providing an excellent home-base for carrying out my PhD project; Esther Leuthold for transcribing interviews; the AMC Research Ethics Committee and the “West Holland Research Ethics Committee” (and their offices) for letting me be part of their daily work on ethical oversight and for allowing me to open the ‘black box’; the contributions to my thesis by people we interviewed on the subject of ethical oversight; researchers who kindly filled out surveys; and anonymous reviewers of manuscripts at the Journal of Medical Ethics, Sociology of Health & Illness, the Journal of Clinical Epidemiology, and the Lancet Oncology.

I thank the New York University School of Medicine and Department of Philosophy for facilitating my working visit to the U.S; Bill Ruddick for being so hospitable during my stay at ‘Washington Square’ (the New York University campus) and for showing me that reaching sixty-five (or seventy-five!) is not a barrier for doing philosophy; the people working on ethical oversight (members and chairs of Research Ethics Committees, institutional monitors and bioethicists) in various U.S. institutions for the stimulating conversations on ethical oversight.

The ‘Joppers’ I thank for exchanging advice and complaints about doing a PhD; my colleagues at the department of General Practice and the Philosophy of Care group, especially Erik B., Debby, Jip, Jolanda, Beate and Susanne; and my roomies at J2-130, Jetstje, Lonneke and Bart: I always thought that we had the best room in the department; thanks for letting me work ánd laugh.

Now is also a good time to thank some of the people who predate my PhD project and who both helped me to get started and/or to get out again.

My Master’s thesis supervisor during my philosophy studies, Wouter Oudemans, for being an inspiring philosopher.

My long-time friends who managed to stick with me during my many years of study and research: the ‘Kabeljauwers’, especially Wiebe, Ward, Harmke, Caroline, Lisa and Esther; the ‘Cobranen’, especially the year ‘97: Michiel, Jeroen, Maarten, Victor, Peter and Fons. You all provided support on a personal level and above all, a great time.
Maarten, close friend since 1997 and paronymph for my thesis defense, I look forward to facing the opposition with you once more.

Pieter, my oldest friend and my other paronymph, thank you for venturing with me into the exciting world of entrepreneurship in science.

My parents, Cees and Yvonne, you have given me so much support in my life, it is truly because of you that I am now here, finishing my PhD. And you provided support in the best way possible: not by interfering, but by being interested.

And finally, my nearest and dearest: Tobias for putting a smile on my face and asking me so many questions, Rebecca for smiling to me so much, and Sabine, my best colleague, favorite conversation partner, and now wonderful wife and mother, for living a good life with me; thank you for always being there for me.
**PORTFOLIO**

Name PhD student: Drs. Jean Philippe de Jong  
PhD period: 2005-2012  
Name PhD supervisor: Prof. Dick L. Willems

**PhD training**

<table>
<thead>
<tr>
<th>Courses</th>
<th>Year</th>
<th>Workload (ECTS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centrum voor Nascholing Amsterdam: Didactics for PhD students</td>
<td>2005</td>
<td>1.5</td>
</tr>
<tr>
<td>AMC Graduate School: Basic course Reference Manager 11</td>
<td>2005</td>
<td>0.1</td>
</tr>
<tr>
<td>AMC Graduate School: The AMC world of science</td>
<td>2005</td>
<td>0.7</td>
</tr>
<tr>
<td>Landelijke Onderzoekschool Ethiek: Ethical theory and moral practice</td>
<td>2006</td>
<td>2</td>
</tr>
<tr>
<td>AMC Graduate School: Better use of PubMed and other medical databases</td>
<td>2006</td>
<td>0.2</td>
</tr>
<tr>
<td>AMC Graduate School: Good and bad, the ethics of medical research</td>
<td>2006</td>
<td>1.5</td>
</tr>
<tr>
<td>AMC Graduate School: Scientific writing in English for publication</td>
<td>2007</td>
<td>1.5</td>
</tr>
<tr>
<td>AMC Graduate School: Entrepreneurship in health and life sciences</td>
<td>2011</td>
<td>1.5</td>
</tr>
</tbody>
</table>
### Conferences

<table>
<thead>
<tr>
<th>Conference</th>
<th>Year</th>
<th>Workload</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Association of Bioethics; 9th World congress of bioethics, Rijeka: poster</td>
<td>2008</td>
<td>1</td>
</tr>
<tr>
<td>Centre for Research Ethics &amp; Bioethics, Uppsala University; Is medical ethics really in the best interest of the patient? Medical ethics conference, Uppsala: presentation</td>
<td>2010</td>
<td>1</td>
</tr>
<tr>
<td>Landelijke Onderzoekschool Ethiek; Yearly conference: presentation</td>
<td>2010</td>
<td>0.5</td>
</tr>
<tr>
<td>Landelijke Onderzoekschool Ethiek; Yearly conference: presentation</td>
<td>2011</td>
<td>0.5</td>
</tr>
<tr>
<td>International Association of Bioethics; 11th World congress of bioethics, Rotterdam: poster</td>
<td>2012</td>
<td>1</td>
</tr>
</tbody>
</table>

### Other

<table>
<thead>
<tr>
<th>Activity</th>
<th>Year</th>
<th>Workload</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMC, Kwalon Journal Club: Qualitative methods</td>
<td>2005-2007</td>
<td>2</td>
</tr>
<tr>
<td>AMC-VU PhD Study Group: Empirical ethics in healthcare</td>
<td>2006-2007</td>
<td>1</td>
</tr>
</tbody>
</table>

### Teaching

#### Year Workload (ECTS)

<table>
<thead>
<tr>
<th>Lecturing</th>
<th>Year</th>
<th>Workload</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMC, Postgraduate education for nursing: Research ethics</td>
<td>2006-2008</td>
<td>0.75</td>
</tr>
<tr>
<td>UvA, Department of Psychology: Research ethics</td>
<td>2006-2009</td>
<td>1</td>
</tr>
<tr>
<td>AMC, Faculty of Medicine: Epistemology, is it possible to know how a patient suffers?</td>
<td>2007,2008</td>
<td>0.5</td>
</tr>
<tr>
<td>AMC, Postgraduate education for midwifery: Research ethics</td>
<td>2009</td>
<td>0.25</td>
</tr>
<tr>
<td>AMC, Faculty of Medicine, course 3.5, illness, behavior and society: Ethics of pharmacogenomics</td>
<td>2009</td>
<td>0.25</td>
</tr>
</tbody>
</table>
**Mentoring**
AMC, Faculty of Medicine, course 20/3.5, illness, behavior and society: Practical; ethics, care and prevention  
2005-2011  1.5

**Supervising**
AMC, Faculty of Medicine, course 12, healthcare in the big city: Research projects  
2006,2007  3
AMC, Faculty of Medicine: Bachelor paper 'Inclusion benefit and medical ethical reviewing of clinical trials'  
2008  1.5
AMC, Faculty of Medicine, course 20/3.5, illness, behavior and society: Writing a paper  
2008,2009  2

**Parameters of Esteem**

**Grants**
Van Walree Fonds: Visiting Scholarship to New York University  
2008
LIST OF PUBLICATIONS


CURRICULUM VITAE

Jean Philippe de Jong was born on December 28th, 1977 in Rotterdam. He received a Master’s degree in Biomedical Sciences in 2004 at Leiden University, in Philosophy in 2005 at Leiden University and in Medicine (with distinction) in 2011 at the University of Amsterdam.

He has worked in the fields of medical research, medical technology and patient care since 2003, specialising in policy, regulation and ethics. Between 2003-2005, he worked on innovative policymaking at the Dutch Ministry of Health, Welfare and Sports and on oversight on medical devices at the Dutch Healthcare Inspectorate.

In 2005, he started working on a PhD project on ethical oversight on medical research with human subjects at the Academic Medical Center in Amsterdam, section Medical Ethics. The results of this project are described in this thesis. He conducted quantitative and qualitative studies of how ethical oversight, especially by Research Ethics Committees, works in practice. Studies conducted in The Netherlands were complemented by a visiting scholarship at New York University, Department of Philosophy in 2008, in order to study oversight in the United States. As a side-project, he developed a guideline for the Netherlands Cancer Institute concerning the storage and use of patients’ residual (tumour) tissue. At the same time as following his PhD training Jean Philippe studied Medicine at the Academic Medical Center in Amsterdam.

Currently, Jean Philippe works at Exon Consultancy (www.exon-consultancy.com), a knowledge-driven consultancy in the fields of pharmaceuticals, medical technology and healthcare. Exon Consultancy was co-founded by Jean Philippe in 2012. Exon Consultancy helps organisations create value from scientific knowledge, gives insight into new scientific developments and helps set up and manage complex,
public-private projects. Exon Consultancy has for example done work for the Dutch Ministry of Health, Welfare and Sports, Escher Projects, World Health Organization, European Federation of Pharmaceutical Industries and Associations, and various companies and universities.

In addition, Jean Philippe is a senior trainer at the Argument Factory - Thinking Academy. This is a private institute that provides insight in complex problems through visualisation techniques and that trains professionals in clear thinking.