THESIS – MASTER APPLIED ETHICS – UTRECHT UNIVERSITY

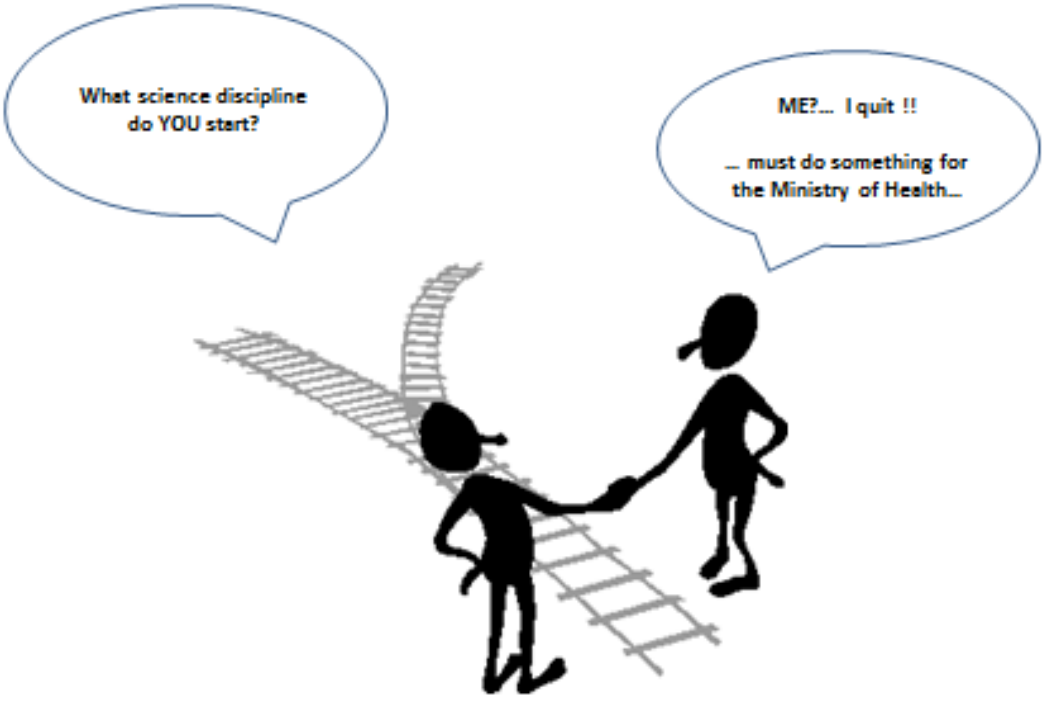
Risk-Benefit Analysis in Clinical Research

Guidelines, Practices and a Utilitarian Interpretation

Reinout Kroes

20 August 2012

This thesis explores several guidelines regarding risk-benefit analysis in clinical research, how some investigators apply these guidelines and concludes that lacunas in both guidance and practice can be observed. Risk-benefit analysis appears to be a difficult task in clinical research. Further, given its necessity for decision-making and its complexity in execution, this thesis explores how we must understand risk-benefit analysis. Because of its natural roots, a discussion is started to come to a utilitarian interpretation of risk-benefit analysis. Two modified features are articulated to meet the practice in clinical research. First, pain and pleasure from classical (hedonistic) utilitarianism are abandoned and replaced by Quality-Adjusted-Life-Years (QALYs). Second, a constraint on the maximum allowable risks is introduced in order to avoid participants from being sacrificed in experimentation. After having presented a most plausible version of a utilitarian interpretation of risk-benefit analysis, a substantial package of criticism remains. This criticism underpins the claim that risk-benefit analysis is a tool under development. If attention is given to some weak points of its procedure, risk-benefit analysis may support the balancing of potential harms and opportunities to some degree. However, it is undesirable to trust risk-benefit analysis as an indisputable calculation model commanding a decision in clinical experimentation by its own force.



RISK-BENEFIT ANALYSIS – know everything about what is probably untrue

Table of Contents

1. Introduction

In the previous decennium problems of doing risk-benefit analysis became once again apparent during the London Drug Trial Catastrophe (2006). In this first-in-man trial a new type of drug was tested. Although compensation for volunteers was far above normal (£2000,-), its preparation procedures complied with all applicable standards. However, at the start-up of testing, all six participants who received the active drug became dangerously ill after being injected TGN1412, an antibody developed to fight autoimmune diseases and leukemia. At recovery, after intensive care treatment, all men faced a lifetime of contracting cancers and all the various autoimmune diseases from lupus erythematosus to multiple sclerosis, from rheumatoid arthritis to chronic fatigue syndrome. The horrible issue is that these men volunteered in joining the research program but did not expect the severe consequences that appeared to be reality. The risk of becoming so severely ill was not communicated with the volunteers beforehand. Moreover, during preparation of the clinical trial this risk was not addressed by the researchers and/or research ethics committee. So it must be concluded that volunteers were unable to take this risk into account in their weighing of participating in the trial. It also must be concluded that either researchers and/or members of the research ethics committee were not aware of this typical risk involved, or that they estimated its chance of occurrence too low and not worth mentioning.

Since The London Drug Trial Catastrophe several studies have been carried out to determine whether the particular risks could have been known and consequently should have been explored in more detail in order to develop and apply sufficient safety measures. These inquiries revealed that there was some doubt about the quality of the results of the trial's risk-benefit analysis (RBA), although conclusions were not univocal. Some argued that the extrapolation of a specific biological mechanism from animal body to human body was not well-founded, relatively

unknown, and should not have been applied without more extensive pre-clinical research. I will not deal with the technical aspects of this case in further detail; these efforts result in applying the same method retrospectively, but then in a more complete and/or precise way, in order to evaluate whether or not risks could have been foreseen and what must be learned from the case. What I do want to explore is how RBA should be approached and how the underlying moral paradigm can be justified on which the practice of clinical research in general and the method of RBA in particular is built. What was not questioned in the referred inquiries after The London Drug Trial Catastrophe was the whole method of RBA as such. So, it seems that this method is still steady as a rock. However, it is very conceivable that during the next RBA to be performed for another clinical trial, again some risk is not addressed because another new (unknown) biological mechanism pops up. So, how can it be that we are still satisfied with a decision method that does not permanently prevent catastrophes? After all, the six men of The London Drug Trial Catastrophe fear severe illnesses related to their defense mechanism for the remaining years of their life.

There are two typical characteristics of clinical research that I think are worthwhile mentioning because they indicate why the matter is important to investigate. First, investigators operate in an environment in which uncertainty is a normal aspect of their work. Uncertainty is an essential variable in clinical research in the sense that it is never certain that a promising hypothesis is proved by particular clinical tests. Otherwise, clinical tests would be not necessary at all. These clinical tests may induce the expected response or may induce no relevant response at all in test subjects, proving that the hypothesis is correct, respectively not correct. Alternatively, the clinical tests may induce an unexpected response in test subjects that also proves that the hypothesis is not correct. At worst, this unexpected response has unfavorable consequences for the test subjects. The root cause of this unexpected response is either a manageable problem, e.g. a human error in the test procedure or a safety measure that was economized on, or a more fundamental

problem, e.g. a biological connection which was not recognized or poorly understood. In my view, it is one of the objectives of RBA to capture both types of root causes so that they can be addressed in the assessment and subsequently proper measures can be arranged. The latter root causes, i.e. the fundamental ones, are of interest because they are the toughest ones to articulate and to estimate. Further, a relation can be imagined with a second typical characteristic of clinical research. The recognition and understanding of biological connections is a most specialized activity that depends amongst other things on the integrity of researchers. Their way of scrutinizing has impact on the quality of hypotheses they propose, being the subject of clinical research, and consequently on the safety of patients as well as test subjects. In clinical research we are stuck with many individuals over whom risks and benefits are unequally distributed. Some individuals bear the burdens of more risk and other individuals take advantage from more benefits. In deciding for clinical trials, investigators must balance the uncertainty of well-being of test subjects and the potentially increased well-being of patients. Phrasing the issue this way might reveal that the challenge is somehow articulated according to a utilitarian interpretation. To what extent this suggestion is sound and how this articulation looks like is of primary concern. Unfavorable consequences for test subjects are so to say in the researcher's hands. But researchers are human beings who are also liable to ambition and temptation which influence their professional behavior. The profession of investigator brings along a special responsibility because of the well-being of patients and test subjects. I believe that RBA can be a proper tool to facilitate giving account of this responsibility.

Therefore, I think, it is important to clarify how RBA is currently used and what justification can be provided for its role. Considering the case and the reasoning so far brings me to the following research question: *How must we understand RBA, given its necessity for decision-making and its complexity in execution?* This research question is divided into an empirical sub-question and an ethical sub-question. The empirical sub-question reads*: How should RBA be performed*

according to current guidelines and how is RBA actually practiced? Because of the possible aggregation over more than one human being as presented above, RBA is in need of a moral foundation which leads to the following ethical sub-question: *How must we understand the utilitarian interpretation of RBA and is such an interpretation defensible?* So, the objective of this thesis is to present a clear understanding of RBA. In more detail, I will make an inquiry of relevant guidelines and its current practice, elaborate a utilitarian interpretation and discuss whether such an interpretation is defensible.

I first address two empirical issues. The first issue deals with guidelines that prescribe and interpret RBA (section 2). The second issue deals with the current practice of RBA and how this is related to the guidelines (section 3). In general terms, these guidelines prescribe that risks in clinical trials may not outweigh potential benefits. They also insist that risks must be reduced to a minimum level. To some extent these guidelines allow that risks and benefits are aggregated over more than one human being; however, they are not clear about how the burdens of research subjects should be balanced against the merits of healthcare and society. This first empirical issue is related to the second, which addresses the way in which RBA is actually performed during the preparation of clinical research, assuming that investigators and review boards comply with the guidelines. I present in general terms how in a few research protocols the risks and benefits are processed and how assessments are articulated that lead to conclusions. After having made a short inventory of the guidelines and a comparison with the practices, I discuss several points of critique.

As RBA sprang from the utilitarian school, it can be expected that utilitarianism may provide support for the method of RBA and the justification of its results. In the preparation of clinical trials, an assessment is made of risks and benefits. The question is how these risks and benefits are balanced. For this balancing, the aggregation of risks and benefits, i.e. applying the utilitarian principle, is a possible

route to follow. The balancing according to this principle suggests that higher levels of risks may be acceptable if the respective benefits are more extensive as well. I make an inventory of utilitarianism and suggest that there is a version of this tradition in which the method of RBA can be smoothly integrated (section4). By doing so, I try to provide the utilitarian interpretation of RBA as attractive as possible in order to show how strong the RBA-tool can be and how much impact it may have on research ethics. But, there are some hurdles ahead to be considered. First, what kind of intrinsic values must the utilitarian aim for? The answer to this question has direct consequences for the kind of risks to be considered in RBA. The classical utilitarian takes intensity, duration, certainty, etc. of pain into consideration in order to calculate maximum utility. But, are pleasures and lack of pain the only value that increase utility (hedonism), or is there another concept that represents lack of harm more profoundly? I will provide an argument why Quality-Adjusted-Life-Years (QALYs) may properly fulfill this duty. Second, a major challenge in this balancing of risks and benefits according to the utilitarian principle is to avoid that risks for a few individuals run too high. Common morality expects that burdens are to be regarded irrespective of the extensiveness of the benefits to many (other) individuals. Because avoiding harm is to some extent more special than relief from suffering, an appropriate constraint would demand that avoiding harm is considered separately. This constraint is derived from Rid et al[1]. They claim that any risk of harm posed by research experiments is to be considered in the light of comparable risks that belong to *daily life standards*. So, the experiment-related possible harms should be compared with and may not exceed the uncertainty of daily-life harms with similar severity. In order to show how RBA can be performed with the proposed QALYs and the constraint requirement, I present a simplified utility calculation model. This or similar model may be helpful in showing how, for example in the London Drug Trial Catastrophe, the total utility outcome as well as the individual risks of harm can be balanced and optimized.

[1] Rid A, Emanuel EJ, Wendler D. Evaluating the risks of clinical research. JAMA 2010;304(13):1472-1479.

Nevertheless, there is criticism that still remains and I discuss four arguments that put the utilitarian interpretation of RBA in perspective (section 5). First, I start from a classical version of utilitarianism and it might be questioned whether more contemporary versions may yield better results. In de version that I propose, some snags may appear in the application of maximum allowable risks derived from the comparable risks that belong to daily life standards. Second, there are some limitations regarding the conceivability of risk scenarios and these limits have an impact on the completeness of RBA. Third, the quantification of probability and uncertainty raises problems that may reduce the value of RBA, if RBA is conceived literally as a calculation model. Fourth, it might be questioned what kind of verdict an RBA must yield. Is this verdict about a state of affairs or about an action, and does this verdict imply a good or a bad thing to strive for (or do) or does it merely imply a decision for the best option of several alternatives?

Based on the criticism I conclude that the use of RBA, based on the utilitarian interpretation, is restricted and that its application cannot be trusted blindly (section 6). I put forward recommendations that may regain its power and assume the status of a proper supporting tool in clinical trial decision-making.

2. Clinical research guidelines

In this section of the thesis I will analyze relevant law and regulations and make an inventory of the prescriptions regarding risks and benefits in clinical research. Therefore, I use the Nuremberg Code (1949), the Declaration of Helsinki (1964 - 2008), the Belmont Report (1979), the International Ethical Guidelines for Biomedical Research Involving Human Subjects (CIOMS – 2002), the ICH-Guideline for Good Clinical Practice (1996) and the Dutch law "Wet Medischwetenschappelijk Onderzoek met mensen (WMO – 2006). Second, I will give an impression of how these aspects are actually implemented in practice. Third I will discuss four points of critique.

Some regulations of the Nuremberg Code:

1. The voluntary consent of the human subject is absolutely essential.
4. The experiment should be so conducted as to avoid all unnecessary physical and mental suffering and injury.
5. No experiment should be conducted where there is an a priori reason to believe that death or disabling injury will occur; except, perhaps, in those experiments where the experimental physicians also serve as subjects.
6. The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment.
9. During the course of the experiment the human subject should be at liberty to bring the experiment to an end if he has reached the physical or mental state where continuation of the experiment seems to him to be impossible.
10. During the course of the experiment the scientist in charge must be prepared to terminate the experiment at any stage, if he has probably cause to believe, in the exercise of the good faith, superior skill and careful judgment required of him that a continuation of the experiment is likely to result in injury, disability, or death to the experimental subject.

The way in which involuntary examinations were carried out on vulnerable and so-called inferior people during the Second World War was the primary reason to set up the Nuremberg Code. Its first regulation addresses therefor the absoluteness of voluntary consent of the human subject to be exercised "in free power of choice without the intervention of any element of force, fraud," etc. Regarding risks and benefits, some regulations are worth to pass in review. Regulation 6 of the code quite literally states that "the degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved in the experiment." Here, the phrase "that determined by the humanitarian importance of the problem to be solved" could be read as the benefit of the experiment.

I now want to indicate why some (part of the) phrased regulations are difficult to interpret or difficult to implement. First, the terms of risk and humanitarian importance are vague and not further explained. Although not explicitly expressed throughout the code it can be assumed that somehow risks concern injury, disability and death. Second and very important, risk may not exceed the humanitarian importance of a particular trial. So, some kind of balancing between risks on the one hand and benefits on the other hand must be performed and that balancing must result in a surplus of benefits. Finally and fortunately, three safeguards have been admitted to the code. Maximum limits have been determined in regulations 4, 9 en 10, stating that unnecessary suffering must be avoided, that the human subject is free to stop the experiment and that the scientist is obliged to stop the experiment when it is likely that injury, disability or death occurs. So, although quite roughly, we find elements in the Nuremberg Code that determine how risks and benefits are related to each other and how risks are related to the judgment of the subject and to the judgment of the scientist. However, it is still not clear how some risks and some benefits may be counterbalanced in such a way that substantial benefit outweighs those risks. It is also not clear what to do when suffering is unavoidable, what the consequences for the subject are of a sudden step out of the experiment and whether a scientist is capable of assessing the potential injury, disability or death. A remarkable regulation is number 5, which states that a massive risk of death or injury is acceptable if the scientist or experimental physician participates in the experiment himself. I guess this very idea stems from a period in which experimenting scientists were conceived as heroes, offering their life for progress in science. These experiments were undoubtedly considered as resulting from virtuous character, may be intended to inspire others to participate in such kamikaze-like activities. A variant of this regulation number 5 is prescribed in the Belmont Report as well as in the International Ethical Guidelines for Biomedical Research Involving Human Subjects, prepared by CIOMS. In these latter two, the requirement is even widened to a broader scope: any human being with manifest voluntariness may

participate in high-risk experimentation “for altruistic reasons or for modest remuneration” (CIOMS: Guideline 8).

The version of 2008 of the Declaration of Helsinki provides for a further shift towards strengthening the protection of human subjects. Regulation A6, for example, states that “in medical research involving human subjects, the well-being of the individual research subject must take precedence over all other interests.” [2]

Some regulations of the Declaration of Helsinki:

A6. In medical research involving human subjects, the well-being of the individual research subject must take precedence over all other interests.
B2. Medical research involving human subjects must conform to generally accepted scientific principles, be based on a thorough knowledge of the scientific literature, other relevant sources of information, and adequate laboratory and, as appropriate, animal experimentation. The welfare of animals used for research must be respected.
B8. Every medical research study involving human subjects must be preceded by careful assessment of predictable risks and burdens to the individuals and communities involved in the research in comparison with foreseeable benefits to them and to other individuals or communities affected by the condition under investigation.
B10. Physicians may not participate in a research study involving human subjects unless they are confident that the risks involved have been adequately assessed and can be satisfactorily managed. Physicians must immediately stop a study when the risks are found to outweigh the potential benefits or when there is conclusive proof of positive and beneficial results.

The notion of well-being is an important one and will be addressed in more detail in section 4. Another new governance of clinical research is its emphasis on the proper use of scientific principles, based on pre-clinical research including literature, laboratory testing and animal experimentation. The results of pre-clinical research must be reflected in RBA. Here, the welfare of animals must be respected (regulation B2). Another refinement regarding RBA is articulated in regulation B8, stating that predictable risks and burdens must be compared with foreseeable benefits to all individuals affected by the investigation. A further new requirement is that physicians may not themselves participate in research study unless risks involved have been adequately assessed and can be satisfactorily managed (regulation B10). Finally, the benefit of scientific knowledge that represents the major part of the clinical research objectives is safeguarded

[2] The notion that the wellbeing of the subjects must take precedence over interests of science and society was already used in the 2000 version; it was not yet used in the 1975 version (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1884510/#b10).

B20. Authors, editors and publishers all have ethical obligations with regard to the publication of the results of research. Authors have a duty to make publicly available the results of their research on human subjects and are accountable for the completeness and accuracy of their reports. They should adhere to accepted guidelines for ethical reporting. Negative and inconclusive as well as positive results should be published or otherwise made publicly available. Sources of funding, institutional affiliations and conflicts of interest should be declared in the publication. Reports of research not in accordance with the principles of this Declaration should not be accepted for publication.

by the requirement "to make publicly available the results of (…) research on human subjects" and to provide for complete and accurate reports (regulation B20).

Compared to the Nuremberg Code (1949) and the Declaration of Helsinki (2008), the Belmont Report (1979) gives some insight into how RBA should actually be performed, resulting in the following extra requirements. First, the Belmont Report gives a short account of the concepts of risk and benefit and "required that risks to subjects be outweighed by the sum of both anticipated benefit to the subject, if any, and the anticipated benefit to society in the form of knowledge to be gained from the research." So, as the practice of clinical research has developed during the fifties, sixties and seventies of the last century, possible direct benefit for the human subjects in clinical research is taken into account. Second, although the Belmont Report recognizes that RBA is difficult and that precise judgments are hard to make, it

Some principles of the Belmont Report: Part C – 2. Assessment of Risks and Benefits.

The Nature and Scope of Risks and Benefits.

…

Risks and benefits of research may affect the individual subjects, the families of the individual subjects, and society at large (or special groups of subjects in society). Previous codes and federal regulations have required that risks to subjects be outweighed by the sum of both the anticipated benefit to the subject, if any, and the anticipated benefit to society if the form of knowledge to be gained from the research. …

The Systematic Assessment of Risks and Benefits.

It is commonly said that benefits and risks must be "balanced" and shown to be "in a favorable ratio," The metaphorical character of these terms draws attention to the difficulty of making precise judgments. Only on rare occasions will quantitative techniques be available for the scrutiny of research protocols. However, the idea of systematic, nonarbitrary analysis of risks and benefits should be emulated insofar as possible. This ideal requires those making decisions about the justifiability of research to be thorough in the accumulation and assessment of information about all aspects of the research, and to consider alternatives systematically. This procedure renders the assessment of research more rigorous and precise, while making communication between review board members and investigators less subject to misinterpretation, misinformation and conflicting judgments.

…

insists that research protocols are analyzed systematically and in a non-arbitrary way so that risks and benefits are provided for as precise as possible. Third, it should address all aspects of the research and even consider alternatives systematically. Fourth, the Belmont Report provides for a role for a review board. It requires that risks and benefits are articulated in such a way that it facilitates proper communication between investigator and the review board and the review board should determine whether the investigator's estimates of risks and benefits are reasonable. Fifth, the Belmont Report requires extra safeguards for vulnerable people. Sixth, it requires that "relevant risks and benefits must be thoroughly arrayed in documents and procedures. Finally, seventh, it requires that information from these documents and procedures is "used in the informed consent process."

…
When vulnerable populations are involved in research, the appropriateness of involving them should itself be demonstrated. A number of variables go into such judgments, including the nature and degree of risk, the condition of the particular population involved, and the nature and level of the anticipated benefits. Relevant risks and benefits must be thoroughly arrayed in documents and procedures used in the informed consent process.
…

Some regulations from the International Ethical Guidelines for Biomedical Research - Guideline 8 - Benefits and risks of study participation:

For all biomedical research involving human subjects, the investigator must ensure that potential benefits and risks are reasonably balanced and risks are minimized.

- Interventions or procedures that hold out the prospect of direct diagnostic, therapeutic or preventive benefit for the individual subject must be justified by the expectation that they will be at least as advantageous to the individual subject, in the light of foreseeable risks and benefits, as any available alternative.
- Risks of such 'beneficial' interventions or procedures must be justified in relation to expected benefits to the individual subject.

Risks of interventions that do not hold out the prospect of direct diagnostic, therapeutic or preventive benefit for the individual must be justified in relation to the expected benefits to society (generalizable knowledge). The risks presented by such interventions must be reasonable in relation to the importance of the knowledge to be gained.

The International Ethical Guidelines for Biomedical Research Involving Human Subjects, prepared by CIOMS, departs from the Declaration of Helsinki. However, it incorporates also an important requirement from the Belmont Report. The CIOMS Guidelines strongly emphasize the distinction between therapeutic and non-therapeutic

interventions. In the first one, higher levels of risks seem to be allowed because participants are in fact patients who already run high risks of illness, disability or death. In their cases extra risks from experimentation may be acceptable because these risks might be relatively small compared to their current risk profile. Potential direct benefit from experimentation may outweigh these extra risks. According to the CIOMS guidelines these beneficial interventions are “at least as advantageous to the individuals concerned, in the light of both risks and benefits, as any available alternative”. For experimentation to be scientific it is necessary that results can be compared to a typical norm or zero-level. Therefore, the CIOMS guidelines provide for additional requirements for minimizing risk associated with participation in randomized controlled trials. For the second category, the non-therapeutic interventions, the CIOMS guidelines do not provide a clearer acceptance level than the previous mentioned guidelines except for the statement that “non-beneficial interventions are assessed differently; they may be justified only by appeal to the knowledge to be gained. In assessing the risks and benefits that a protocol presents to a population, it is appropriate to consider the harm that could result from forgoing the research” (Guideline 8).

The Dutch law “Wet Medischwetenschappelijk Onderzoek met mensen” and the ICH-Guideline for Good Clinical Practice do not provide more extended regulations regarding RBA than (part of) the guidelines that are approximately articulated in the Declaration of Helsinki. The ICH-Guideline for Good Clinical Practice very clearly takes the Declaration of Helsinki as a starting point. This guideline uses the term “non-therapeutic trial” in relation to RBA, but it does not define the concept of the term. The term “therapeutic trial” is not used, nor defined.

Regarding the question how RBA should be performed according to current guidelines I address some points for discussion that can be drawn from this brief survey. As indicated, these guidelines have developed over time and their articulation regarding RBA has improved to some extent, resulting in more refined

prescriptions. However, it is questionable whether this evolvement has now lead to sufficient clarity because guidelines hardly describe in detail how RBA can or should be done. Only in general terms, these guidelines prescribe that risks in clinical trials may not outweigh potential benefits. But for those who have to apply the regulations, it is not clear how much weight has to be ascribed to risks and how much weight has to be ascribed to benefits and to what extent these weights are related to the number of individuals affected by those risks and benefits. So, it is not clear how many individuals must carry the weight of a burden of risk and how many individuals must be permitted to enjoy the expected benefits and whether every individual is equal in this balancing. Further, guidelines do not make a distinction between benefits that will certainly be materialized and potential benefits. Benefits may have a provisional character because in some cases they are dependent on other, biological, contractual, economic or political mechanisms to become effective. So, it would be more to the point to use the concept of *opportunity* (or a similar term) as counterpart of risk. Some benefits are accompanied by high levels of probability; these benefits will in the end very plausibly be realized. In these cases the weighing of risks and benefits is less problematic. However, if benefits are accompanied by lower levels of probability, the uncertainty that these benefits will in the end be realized is problematic and in some way should be accounted for in further weighing. For example, it is not sure whether a new therapy will be effective and the weighing is flawed when this uncertainty is not taken into consideration. An unfair balancing might give preference to the wrongful surplus of benefits. I think it is more appropriate to conceive the concept of opportunity as the counterpart of risk and to conceive the concept of benefit as the counterpart of harm. Finally, in the guidelines it is not explicitly stated who actually has to perform the RBA. It seems taken for granted that the investigator as well as the members of the ethical review board have a role in RBA, but it is unclear whether both institutes must perform the whole assessment or, if not, which part or which perspective of the RBA is assigned to which party.

This is, in my view, how guidelines present and prescribe RBA and how, to some extent, it could be argued that still some vagueness results from their prescriptions. When prescriptions are relatively general, it is of interest how RBA practitioners interpret and apply the guidelines. In the next section I will discuss how RBA is actually practiced and what peculiarities can be observed.

3. RBA in practice

After having given an indication of the relevant guidelines and their typical articulations that investigators and review boards have at their disposal, I now want to give an impression of how some researchers apply RBA. In particular, I want to indicate how risks and benefits were estimated and captured in the documents that were placed at the government review board's disposal for review during session 152, January 2012 (CCMO – The Hague). I was allowed to join this session in order to witness the research reviews.

I first briefly describe the relevant institutes in the Netherlands and the review process in which RBA plays a role. The government review board supervises approximately 20 regional Research Ethics Committees (RECs). This supervision contains random supervising review of individual research protocols, treated by RECs (analyzing the REC's review procedure), as well as inspection of REC's procedure management. Both CCMO and RECs deal with clinical research protocols. Investigators are free to choose by which REC they want their research protocols to be review. However, RECs deal with research projects with normal potential impact and the CCMO deals with projects with higher potential impact and investigators much approach the CCMO for review of their projects of the latter category. Examples of projects with higher potential impact are experiments concerning cell-therapy, gene-therapy, vaccination and projects with minors and with subjects unable to give informed consent.

For communication between investigator and review board (CCMO or REC) three types of documents are of interest: the research protocol, the subject information letter and the review board form ("ABR-formulier")[3]. These documents are used by the CCMO as well as by the RECs and serve as a basis for the assessment of the

[3] For simplicity, I leave the documents regarding product information out of consideration.

RBA. In either case the research protocol, the subject information letter and the review board form are prepared by the investigator. A research protocol is a carefully structured, written plan in order to ensure its smooth running and successful conclusion and also to gain the compulsory agreement of an ethical committee. It consists of general information, an introduction, study objectives, selection and withdrawal of subjects, study procedures and treatments, efficacy-, safety- and feasibility assessments, statistical considerations and supporting bibliography. In its efficacy assessment a description is provided of the primary efficacy endpoint of the study, i.e. the final aim of the therapy/drug and by some secondary efficacy endpoints of the study, i.e. typical ways (measurement methods and criteria) by which the aim of the therapy/drug is established. This safety assessment considers how adverse events are defined, classified and handled during a predefined period of time related to the research project. Also part of this safety assessment is a description of "known and expected" risks including safety measures by which risks are reduced to acceptable levels. An information letter, intended to inform potential participants, presents some of the risks and benefits articulated in the research protocol. In general, benefits and risks are transferred from research protocol to information letter and their articulations are stripped of its most scientific and technical terms. All the risks and benefits are also transferred by the investigator from the research protocol to a review board form ("ABR-formulier"). This form structurally helps the investigator to verify that all necessary information and documents are prepared. The review board form requires that the investigator gives a final assessment regarding RBA:

> "Provide, on grounds of your own considerations, why the execution of the study is justified, regarding the burdens and / or risks for the subsequent participants."[4]

So, regulations require the investigator to weigh risks and benefits, but the review board will also make a judgment of its own. The review board discusses the relevant

[4] Section E9a of the review board form ("ABR-formulier")

risks and benefits, provides arguments for weighing and comes to a conclusion. This process of weighing and reasoning is chronologically reported in minutes of meeting and is accompanied by critical remarks that are to be considered by the investigator. The investigator is informed about the critical remarks and asked to provide for additional measures, scientific data and/or other information. A new version of the research protocol, information letter and review board form is provided by the investigator in the following session of the review board and this cycle may be repeated a few times until the review board either approves or disapproves the research protocol for execution.

After having presented a general picture of the process in which RBA is accomplished, I now want to give an impression of how RBA was performed by researchers who prepared the clinical research protocols for review board session 152. During this session 11 research protocols were considered. From the 11 research protocols 5 were considered to have high potential impact and were directly submitted to the CCMO for review; 6 research protocols were considered normal potential impact and had already been treated by a particular REC. These latter research protocols were randomly retrieved by the CCMO for supervision (re-review).

For a survey of these 11 research protocols including the RBA as done by the researchers, I proceed in three steps. This approach is based on a framework derived from a risk management guideline[5]. First, I report whether *risks* are identified, estimated and evaluated. Second, I report whether *opportunities* are identified, estimated and evaluated. Third, I report whether *weighing* and *reasoning*

[5] I choose a current risk management guideline from a connected branch, i.e. pharmaceutical quality, and use the prescribed risk management concepts. According to this guideline, risk assessment is divided into risk identification, risk estimation and risk evaluation. Risk identification is a way of finding and listing of risks; risk estimation is a way of valuing or rating of risks; and risk evaluation is a way of determining the acceptability of risks including additional safety measures (if necessary). This breakdown is presented in the ICH guideline: International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use, ICH harmonised tripartite guideline, Quality risk management, Q9, step 4, 2005. The same breakdown is applied as an analytic technique for risks assessment by Beauchamp T.L., Childress J.F., Principles of Biomedical Ethics, Oxford University Press, 2009, p.224

are performed that can support conclusions. Table 1 (below) presents an overview of the typical survey I prepared. For general information I registered the field of experiment, potential impact, phase of clinical trial and whether the experiment was therapeutic or non-therapeutic. For presenting the *identification* of risks and opportunities, I counted the number of risks and the number of opportunities articulated in each protocol. For presenting the risk *estimation*, I determined per research project whether the investigator made a distinction between the chance of occurrence and the severity of an event (a risk is estimated by the product of these two different notions). Likewise, I determined of the articulated opportunities whether a distinction was made between the chance of success and the benefit of realization. For presenting the risk *evaluation*, I determined per research project whether the investigator made judgments that the isolated risks are acceptable, indicating that sufficient safety measures were in place to prevent events to occur. Likewise, I determined whether judgments were made about the acceptability of isolated opportunities, indicating that sufficient enhancement was provided for benefits to be realized. Further, for presenting per research project the *weighing* of risks and opportunities, I determined whether the investigator provided a statement that the opportunities in fact outweigh the risks. Finally, for presenting some force of conviction that conclusions are acceptable, I presented whether the investigator provided one or more *reasons* for that weighing.

Table 1: Overview of RBA in 11 research protocols

Clinical research project				Risks				Opportunities				RBA	
Field of experiment	Potential impact	Phase	Therapeutic	Risks identified	Risk estimation		Risk evaluation	Opportunities identified	Opportunity estimation		Opportunity evaluation	Weighing opport. versus risks	Reason for conclusion
					chance of occurrence	severity of event			chance of succes	benefit of realization			
cell therapy	high	II	yes	1	yes	yes	no	1	no	yes	no	yes	no
gamete and embryos	high	n.a.	yes	3	yes	yes	no	2	yes	yes	no	no	no
gamete and embryos	high	n.a.	no	0	n.a.	n.a.	no	1	no	yes	no	no	no
minors	high	n.a.	no	1	yes	yes	no	1	no	yes	no	yes	no
minors	high	I	no	3	no	yes	no	1	no	yes	no	yes	no
common therapy	low	III	yes	2	yes	yes	no	1	no	yes	no	no	no
common therapy	low	II	yes	3	no	yes	no	0	no	no	no	no	no
common therapy	low	IV	yes	1	no	yes	no	1	no	yes	no	no	no
common therapy	low	n.a.	yes	0	n.a.	n.a.	no	1	no	yes	no	no	no
common therapy	low	III	yes	1	yes	yes	no	1	no	yes	no	no	no
common therapy	low	III	yes	2	yes	yes	no	1	no	yes	no	no	no

This short survey illustrates what aspects of risk management are considered by the investigators. However, it should be emphasized that there are limitations of this survey. It must be noticed that all 11 research protocols were submitted to the review board for evaluation and approval. So, none of these research protocols were final versions, ready for execution yet. Further, the surveyed research protocols were prepared by investigators, so nothing can be concluded about the way RBA is performed by the review board.

Based on the survey, I want to put forward the following critical observations and conclusions. Concerning the estimation of risk, in most of the cases the investigator prepares a description of severity of event and only in some cases prepares a description of chance of occurrence. Concerning the estimation of opportunity, in most of the cases the investigator prepares a description of benefit of realization and only in one case prepares a description of chance of success. What can be

observed is that in a few cases risk is not properly estimated because chances of occurrence are not considered and in almost all cases opportunity is not properly estimated because the chances of success are not considered. Further, what is not addressed at all in the research protocols is a proper evaluation of individual risks and opportunities, in other words, to reflect on whether risks are minimized and opportunities are maximized. Finally, regarding weighing of opportunities versus risks, it can be observed that only in a few cases (3) a conclusion was provided with an explicit statement that opportunities outweigh the risks. In these 3 cases either chances of occurrence or chances of success (or both) were not considered. In my view, if one of these components is missing upon which risk and/or opportunity is built, the weighing of these risks and opportunities is meaningless. Further, in none of these cases reasons were provided supporting the RBA decision. So, what can be concluded from the observations is that in the 11 clinical research projects the investigators somehow fail to perform RBA in a structured way. There are no evaluations carried out, stating that risks are minimized and opportunities maximized, and there are no reasons provided why benefits outweigh the risks and why it must be concluded that the execution of the study is justified.

The question might be raised whether the framework of risk management that I proposed has substantial added value. This framework contains three concepts of risk identification, risk estimation and risk evaluation, including a procedure to apply these concepts in a structured sequence. In considering the London Drug Trial Catastrophe, it can be claimed that some added value can be expected. When performing a structured risk evaluation, all relevant scenarios and measures are passed in review in order to minimize risks. Probably, the questionable decision of injecting the active test drug into all six participants *simultaneously* would have been detected at forehand. A relatively easy measure to implement was to inject the active test drug successively with sufficient interval and evaluate the response of the first participant. Further, when performing a structured opportunity evaluation, all relevant scenarios and measures are passed in review in order to maximize benefit.

Probably, the questionable decision of not forcing the investigator (by contract or otherwise) to disclose relevant scientific data after the TGN1412-study would have been detected at forehand as well. Disclosure of relevant scientific data is a deserved benefit for other scientists, making them able to further develop a therapy that can fight autoimmune diseases and leukemia. However, the bankrupted company that carried out the TGN1412-study did not do so[6]. In other words, even the small amount of benefit (scientific knowledge) that could have been obtained from the London Drug Trial Catastrophe and related studies was not realized. Proper evaluation of opportunity would have detected this weakness of the project and additional procedures could have been demanded in order to boost the chance of success.

Although I have focused on the investigators and the information they provide concerning RBA, another question might be raised whether review boards receive adequate and enough information to perform a proper review. If any review board wants to pick up the RBA reasoning somewhere from the research protocol, the board members have a hard job to do. If risks and/or opportunities are not properly listed and estimated, weighing is a precarious task. Moreover, if a review board wants to review / perform an RBA constructively, the review board encounters a dilemma. The review board is obligated either to propose the optimizations of risks and/or opportunities by its own initiative (in fact to participate in re-designing the research project), or to reject the research protocol and bring about substantial delay of clinical research.

This is, in my view, how RBA is actually practiced in clinical research environment. In the next sections I will discuss how utilitarianism would deal with harms and benefits and what difficulties appear with a utilitarian interpretation of RBA.

[6] This is the final objection of Kenter M J H, Cohen A F, Establishing risk of human experimentation with drugs: lessons from TGN1412, The Lancet, Vol 368, 2006. In this article the authors express their frustration of the investigator's scientific data not being available regarding pre-clinical research and research that was related to the TGN1412-study.

4. Utilitarian interpretation

In this section I discuss an ethical theory that can be set up for a proper model of weighing risks and benefits. This theory is amongst other things concerned with maximizing the utility of actions. When reading the London Drug Trial Catastrophe case according to a simplified model, it could be said that release from leukemia is a *benefit* and consequently, that the severity of contracting auto-immune diseases is considered as *harm.* If taken for granted that actions will certainly achieve the predicted benefits and harms, the model for decision-making can be simplified to the following equation:

Utility of activity = benefit – harm

In this model for decision-making it should be noticed that its variables (benefit, harm) serve the wellbeing of individuals. One influential theory in ethics – hedonism – conceives wellbeing in terms of pleasure and pain. But how can pleasure and pain account for ethical decision-making? A utilitarian would argue that release from leukemia can be grasped with the concept of *pleasure* and the contracting auto-immune diseases with the concept of *pain*. Here, we arrive at a classical view of hedonistic utilitarianism, submitted by its originators Jeremy Bentham and John Stuart Mill. According to Timmons, hedonistic utilitarianism (HU) is articulated as followed[7]:

> "HU The utility of an action = the overall balance of pleasure versus pain that would be produced were the action to be performed."

Typical features of this decision-model are that pleasure is considered as the absence of pain, that pain can be determined with features like intensity, duration,

[7] Timmons M., Moral Theory, An Introduction (Lanham: Rowman & Littlefield, 2002), p.107

extent, etc. (Bentham's felicific calculus[8]), that every individual counts for one and no one counts for more than one and that this decision-model is applied to maximize utility. An action is right if it brings forth at least as much utility, i.e. more pleasure than pain, as any other action.

Regarding the difficulty of grasping the *value* of benefits and *severity* of harms Bentham tries to solve the issue by assuming that pain is conceived as counterpart of pleasure and that the intensity, duration, etc. of pain are taken into consideration in order to calculate maximum utility. However, I doubt whether maximum utility can be determined by considering only pain and the lack of it. To be disabled or to have an unnatural limited lifespan are examples of situation in which harm does not necessarily involve pain, although it reduces the possibilities for experiencing pleasure. Assessing such harms and benefits is complex and it could be argued that pleasure is only one element of a set of benefits and pain is only one element of a set of harms. Beauchamp & Childress suggest that instead of using informal techniques such as expert judgments, more effective techniques may be used that employ quantitative analysis of benefits and harms[9]. One way of doing this is by determining the value of benefits and severity of harms in terms of Quality-Adjusted-Life-Years (QALYs). According to Beauchamp & Childress, an important feature of "QALYs is that, if an extra year of healthy (i.e., good quality) life-expectancy is worth one, then an extra year of unhealthy (i.e., poor quality) life-expectancy must be worth less than one…"[10] An argument for application of QALYs is that it assigns bodily functions as important parameters too. In my view, the application of QALYs is a more broad and sophisticated way and a useable approach of determining the value of benefits and severity of harms than measuring the intensity, duration, etc. of pain, as advocated by Bentham. A prerequisite is that prudent physicians properly classify different levels of harm including pain, disabilities and unnatural limited lifespan and to provide for reliable figures representing unhealthy (i.e., poor quality)

[8] Bentham J, Introduction to the Principles of Morals and Legislation, 1789, section 4.2
[9] Beauchamp T.L., Childress J.F., Principles of Biomedical Ethics, Oxford University Press, 2009, p.221
[10] Idem, p.231

life-expectancy. Such a classification of QALYs would operate in the same way as the classical classification of pain; a surplus of gain of QALYs for one set of individuals compared to the loss of QALYs for another set of individuals would justify the subsequent action. Application of QALYs is attractive because it roughly quantifies the total amount of years to gain (or lose), adjusted by a quality coefficient in the current state of life or expected state of life in the future. A disadvantage is that, as applied in the utility calculation model, it does not clarify how life-years (gained or lost) are distributed among participants or patients. This could be solved by preparing separate calculation sheets dedicated to particular age categories. Further, it could be suggested that the utility calculation model discriminates older people. Because the normal life expectancy is naturally lower than the life expectancy of younger people, their opportunity estimation will turn out lower and this results in lower expected utility. Diminished expected utility may lead to less clinical research projects on behalf of the elderly.

Timmons puts forward that there are three intuitive plausible ideas behind utilitarianism. First, it is appealing that this theory is concerned with the well-being of humans, making use of the typical value as operator: the absence of pain (or the gain of QALY). Second, it seems plausible to say that an action is most praiseworthy if it brings about the most good in the world. Third, it also seems convincing that the theory embraces the idea of impartiality and universality: all individuals who are affected by an action are counted as morally relevant and as equal (for equal individuals). The method of determining maximum utility has strong appeal because it can be applied easily on individual level and its structural approach facilitates decision-making in more complex circumstances. A utilitarian would argue that this determination of maximum utility can also be applied to a group of individuals. Then, benefits and harms are considered as aggregated over all individuals of the group and consequently an action is chosen that yields the maximum outcome for that group. This aggregation will be discussed below in further detail.

So far, it could be said that utilitarianism was interpreted as *actual utilitarianism*, applied as a moral criterion to evaluate actual and known state of affairs. Timmons argues however that a moral theory providing only moral criteria is not complying with the standard of applicability[11]. A moral theory must also serve as a decision procedure. He suggests that *probable utilitarianism* meets this applicability standard. Its aim is focused on expected utility rather than actual utility. Core notion of this version of utilitarianism is that, if the final state of affairs is suboptimal, meaning that an outcome of which the utility is lower but the probability higher than a more optimal alternative (more utility, but less probable), the decision-maker has done a proper job and is not to blame[12]. In other words, in spite of the suboptimal outcome of state of affairs, the decision was properly taken. Timmons doubts whether this version of the theory ought to be used in every situation. He provides the example of considering saving a child from being run over by some vehicle on the road and argues that in such a case the proposed way of deliberation is not useful. In our ordinary lives we naturally weigh risks and opportunities, but we rarely apply it in a formal way. For some difficult decisions nevertheless, we may use the method for structured consideration and to come to a reasonable conclusion. As I conceive the weighing procedure, such a decision is facilitated by the following calculation model:

Expected utility of activity = opportunities – risks

This equation deviates from the original RBA notion (expected utility of activity = benefits – risks) in the sense that benefits are replaced by *opportunities* because it is not always certain that benefits become effective. As already indicated in the previous sections, the concept of opportunity is a product of the probability that a benefit becomes effective (chance of success) and the value of that benefit (benefit of realization); and the concept of risk is a product of the probability of a harm (chance of occurrence) and the severity of such a harm (severity of event).

[11] Timmons M., Moral Theory, An Introduction, Lanham: Rowman & Littlefield, 2002, p.122
[12] Idem, p.126

From here, I bring along the explained components and apply the model for decision-making in more detail to the London Drug Trial Catastrophe:

> Expected utility of London Drug Trial = (probability of available and effective leukemia therapy x the value of release from leukemia x the number of patients) – (probability of subjects contracting auto-immune diseases x severity of contracting auto-immune diseases x the number of subjects).

Note that this is still a simplified presentation, because there might have been more risks relevant for weighing and decision-making. A question could be raised whether this equation must be conceived literally or symbolically. If it is conceived literally, it will lead to the following calculation, presented in table 2. This calculation sheet is a simply applied decision model, similar to examples presented by Timmons[13]. It uses imaginary figures to calculate the total opportunity of gained QALYs, the total risk of lost QALYs and determines utility by establishing the surplus of gained QALYs. The idea is that for calculations of expected utility of alternatives, separate calculation sheets are used. The maximum found expected utility figure determines which alternative is the best option. It must be noticed that abandoning the trial is always one of the options and considered as starting-point of weighing the current situation.

Table 2: Utility calculation sheet

Item	Sub-item	Unit	Figures & calculations			
Opportunity	chance of success	%	10,0%			
	benefit of realization (QALY 1 → 2)	gained QALY	50	x		
	opportunity estimation	opp. of gained QALY	→	5		
	persons affected	quantity			6.000	x
	total opportunity	opp. of gained QALY			→	30.000
Risk	chance of harm A (comparator: 0,01%)	%	0,05%			
	severity of harm A (QALY 1 → 2)	lost QALY	20	x		
	risk estimation A	risk of lost QALY	→	0,01		
	persons affected	quantity			6	x
	total risk	risk of lost QALY			→	0,1
Utility (total opp. – total risk > 0)		surplus of aggr. QALY				**29.999,9**

[13] Timmons M., Moral Theory, An Introduction, Lanham: Rowman & Littlefield, 2002, p.122 and further.

As input, the calculation sheet asks for probability figures and for gained / lost QALYs. Gained and lost QALYs are determined by the surplus between expected QALY of state of affairs at point 1 (in time) and expected QALY of state of affairs at point 2. Point 1 in time is related to the state of affairs before an action; point 2 in time is related to the state of affairs after the action has been completed. Opportunities result from positive (gained) QALYs if expected QALYs of 2 are higher than QALYs of 1; risks result from negative (lost) QALYs if expected QALYs of 2 are lower than QALYs of 1. The found utility (surplus of aggregated QALY) must be maximized and must be positive (table 2: cell-area appears green).

By conceiving its equation in a literal way table 2 shows how the decision-making model of the London Drug Trial Catastrophe turns into an odd case. With these hypothetical figures, utility appears very promising. But the risk of harm to some individuals is far above normal (its chance is 5 times the level of daily life). This calculus reflects that, in spite of the six men facing severe illnesses for the remaining years of their lives, the gathering of knowledge from this experiment can be crucial and can contribute significantly to the development of new therapy for combatting leukemia. In this hypothetical and literal calculation the new therapy lengthened the life of many otherwise very ill and dying individuals and the sacrificing of six lives is facilitated and justified by the model. For this problem to occur, it is presupposed that in the opportunity component of the equation, opportunity of many individuals can be added up resulting in one big opportunity for the group of individuals. The equation seems to allow a decision to put a few subjects at risk of substantial harm for the benefit to many. Utilitarianism may ‘bite the bullet’ and insist that this kind of experimentation is acceptable. However, many people might reasonably reject such a conclusion; it is contrary to our moral intuitions. Utilitarianism must somehow deal with this problem and is in need of some kind of constraint that prevents such an odd case to be the result of the calculus process.

An appropriate constraint would be that the probable harms are considered separately. Rid et al present a plausible view how these limits of harm can be determined in clinical research[14]. They suggest to categorize harms that belong to daily life standards and to determine accompanying uncertainty-levels. Rid et al provide seven levels with expanding magnitudes of harm and accompanying illustrative examples (negligible, small, moderate, significant, major, severe and catastrophic). Examples of significant harm are ligament tear of the knee with permanent instability and intensive care treatment for several weeks; examples of catastrophic harms are dementia and death. Rid et al claim that any risk of harm caused by research experiments is to be considered in light of comparable risks that belong to daily life standards. So, these experiment-related possible harms should be compared to the probability of daily-life harms with similar severity (table 2: "comparator", cell-area appears red). The constraint of the experiment is that separate risks of harm in testing may not exceed the risks of daily-life harms. This approach has two appealing advantages. First, in non-therapeutic experimentation it protects participants from excessive risks. Second, in therapeutic experimentation it may allow patients, who are exposed to higher levels of risks that belong to daily life standards of patients, to participate in clinical trials with higher risks of harm that may at the same time yield a direct benefit for the patients themselves.

So far, I presented RBA as a kind of natural fit in utilitarianism. In order to make this utilitarian interpretation of RBA plausible, I put forward that there are two cardinal challenges to be met. For applying values and levels of severity I suggest to reject the hedonistic parameters of pleasure and pain from the classical utilitarian view and to accept the concept of QALYs. With the application of QALYs, the comparison of risks and benefits between (groups of) individuals is more complete, although the estimation of QALYs is not without difficulty. For the sacrifice problem I suggest to determine additional constraints regarding the maximum levels of chances of harm that are to be compared with daily life standards in order to prevent sacrificing a few

[14] Rid A, Emanuel EJ, Wendler D. Evaluating the risks of clinical research. JAMA 2010;304(13):1472-1479

individuals for the benefit of the majority. In the next section I will criticize the utilitarian interpretation of RBA.

5. Criticism

I have tried to provide the utilitarian interpretation of RBA as attractive as possible in order to show how strong the RBA-tool can be and how much impact it may have on research ethics. From here, nevertheless, I want to put forward four discussion points that put RBA according to the utilitarian interpretation in perspective. First, I started from a classical version of utilitarianism and it might be questioned whether more contemporary versions may yield better results. In de version that I proposed, some snags may appear in the application of maximum allowable risks derived from the comparable risks that belong to daily life standards. Second, there are some limitations regarding the conceivability of risk scenarios and these limits have an impact on the completeness of RBA. Third, the quantification of probability and uncertainty raises problems that may reduce the value of RBA, if RBA is conceived literally as a calculation model. Fourth, it might be questioned what kind of verdict an RBA must yield. Is this verdict about a state of affairs or about an action, and does this verdict imply a good or a bad thing to strive for (or do) or does it merely imply a decision for the best option of several alternatives?

One of the main objections to classical utilitarianism is the claim that it justifies putting a few individuals at risk (sacrifice) for the benefit to many. It could be argued that classical utilitarianism can deal with this phenomenon on its own: the realization of such an odd case itself would be enough for potential subjects to realize that participating in clinical trials is equal to signing a death warrant. In the long run, taking many sequential experiments in scope, maximizing utility would demand designers of experiments to take into account that enough participants for experiments are available and this might primarily be realized by preventing subjects from being sacrificed and by inspiring confidence, resulting in the willingness of subjects to participate. However plausible this view seems to be, I doubt whether a constraint in this form is sufficiently concrete and practically feasible in the clinical

research environment. It would require that investigators and review boards take into account the risks of their own experiment and the opportunities of potential other, yet unknown experiments. Because of the impracticality in the use for the assessment of one experiment alone, this constraint must prevail at forehand. In the previous section I suggested to put in place additional requirements that may cope with this problem: besides the utility calculation, all probable harms have to be considered separately and their accompanying risks may not exceed comparable risks that belong to daily life standards. By doing so I changed from the classical utilitarian starting point to a more contemporary version of utilitarianism that accounts in a way for respect of persons. Timmons suggests that this can be found in pluralistic utilitarianism. According to pluralistic utilitarianism "there are moral constraints on valuable projects; … projects must among other things respect other people at least in the minimal sense of not causing significant and avoidable harm" [15]. Regarding the constraint of considering probable harms separate from the total utility of experimentation, some snags may appear in the application of comparable risks that belong to daily life standards. A first objection is that experimentation will always *add* some risk to the daily life standard. So, how can the risk for a subject who is confronted with risks of daily life standard, travelling to and from a clinic, and who is also confronted with the additional risk of a clinical trial, meet the comparable risks that belong to the daily life standard? When calculations are performed exactly and limits are applied precisely, this is logically impossible because there is always additional risk involved. If this comparison is performed not precisely but approximately, how much additional risk is allowed? RBA based on utilitarianism that respects the wellbeing of individuals cannot answer that. May individuals decide for themselves? If not, is that respectful? In my view there is no satisfying justification for the levels of daily life standards of risks and for the application of these levels without consulting the respective subjects. A second objection is that the daily life standard may vary between different societies. Its application may yield that experiments unacceptable in western society because of a high risk profile might be

[15] Timmons M., Moral Theory, An Introduction (Lanham: Rowman & Littlefield, 2002), p.145

acceptable in developing countries where daily life standards contain higher levels of normal risks. So, applying constraints that are derived from daily life standards may impose inequality between societies.

As a second discussion point, regarding the limitations of conceivable risk scenarios, it is worth indicating how these limits have an impact on the completeness of RBA. In doing risk inventory – normally exercised in brainstorming sessions – all kinds of events with bad consequences can be imagined. Underlying biological principles that account for bad consequences to occur may exist but are unknown, or are at best presumed. Or, underlying principles may not exist at all and are still presumed. In the first case (potential harm exists but is unknown), entry into risk inventory is obligatory but problematic because its quantification is difficult. In the second case (potential harm does not exist but is presumed nevertheless), entry into risk inventory is a mistake because it disturbs the reliability of the outcome of the risk analysis. The unjust risk may lead to the introduction of an unnecessary safety measure (which may introduce even new risks) or may lead to an unjust decision of the review board to discontinue the experiment. So, we are dealing with the problem how to discern realistic potential bad consequences from unrealistic potential bad consequences and it is a challenge to determine the uncertainty of potential bad consequences. Realistic potential bad consequences with sufficient uncertainty are candidates for queuing up in the risk-inventory list to be assessed in RBA. Regarding the London Drug Trial Catastrophe, Kenter and Cohen made an analysis of what was known at the time of the TGN1412-study about the so-called CD28 antibody, used as test-drug. This super agonistic antibody was a new compound with a complex and novel mechanism and Kenter and Cohen noticed that “the TGN1412 study was the first trial of this type of compound that was undertaken in man, so only a small amount of human data were available for risk analysis.”[16] Nevertheless, Kenter and Cohen concluded that substantial knowledge was

[16] Kenter M J H, Cohen A F, Establishing risk of human experimentation with drugs: lessons from TGN1412, The Lancet, Vol 368, 2006, p.1

available, e.g. about similar antibody reactions in humans and relevant mechanisms in animals. This knowledge could support for conceiving and taking serious the complex and novel mechanism that lead to the fatal immune reaction in human bodies. They argued that their "risk analysis, undertaken with data available in the research file and public domain before the TGN1412 trial started, shows that essential information was absent and the antibody was a high-risk compound unlikely to be suitable for administration to healthy people without additional preclinical experiments."[17] Moreover, Kenter and Cohen emphasize that other scientific data from the investigator, data that could be used for a better understanding of the complex and novel mechanism, was not made available to the scientific community. This scientific data could have been valuable for the preparation of new clinical research; however it was not released before, neither after the investigator's bankruptcy. A further worry of Kenter and Cohen was about the approach of how potential harms are listed and structured. They suggest that "different people who assess risk of a human study should communicate their findings in a consistent and orderly manner to boost the chance that the right questions are asked."[18] Kenter and Cohen provide a standard list of questions, meant as a reminder for clinical research addressing issues in a risk analysis of a new compound. However, Kenter and Cohen were most concerned about the amount and completeness of information that was provided by the investigator to the review board as well as to the scientific community. This worry was about the necessary condition that scientific data is available, because availability is crucial for researchers to be able to thoroughly survey all potential harm of their planned experiments. So, it can be suggested that identifying all potential harms and accompanying risks is a painstaking job, requiring much expertise and integrity (openness) in the discipline. Part of this expertise is that investigators are literally expected to thoroughly search for possible risks. Any failure of compliance with

[17] Idem, p.4
[18] Idem, p.4

these requirements will reduce the value of an RBA and its utilitarian justification of the accompanying experiment.

Regarding the third discussion, it is often not clear how probability- and uncertainty aspects of benefits and harms can be quantified in a proper way. Probabilities of benefits somehow coincide with the aim of the experiment, i.e. to determine to what extent the hypothesis under investigation is true. For example, to determine the chance of success of the TGN1412-study, it is crucial to know to what degree a so-called CD28 antibody is able to fight autoimmune diseases and leukemia and this is exactly a not-yet-known figure. Further, the success of final benefits is often dependent on the state of affairs outside the investigator's influence. There might be other burdens to overcome regarding licensing, financing and production of particular companies in order to obtain a proper medicine. Finally, it is conceivable that experimentation is proposed for further extension of scientific knowledge that might lead to the development of important medicines. If the first step of this train is considered in isolation, a direct benefit from experimentation in terms of gained QALYs cannot be expected. When participants are posed to some risk of lost QALYs, the utility calculation model will result in a negative outcome, providing a reason not to approve the first step of an experiment train. So, it is not clear whether the utility calculation model can be used considering some experiments in isolation. If so, it is still not clear to what extent an experiment will contribute to the development of important medicines. Likewise it is often not clear how uncertainties of harm can be quantified. In considering epidemiological evidence in the London Drug Trial Catastrophe it would have been reasonable, as suggested above, to investigate more thoroughly the probabilistic connection between administering a super agonistic antibody and the adverse event of fatal immune reaction. The investigator should have surveyed similar probabilistic relations in other experiments. However, if no probabilistic data are available at all, calculating its corresponding risk is not possible either.

In his article 'Risk and Precaution' Stephen John suggests that it should be possible quantifying the uncertainty of unknown and presumed connections between events that may cause harm in a more profound way. Scientific claims can be divided into claims with a null-hypothesis and claims with an alternative-hypothesis. The null-hypothesis would be that there is no connection between administering a super agonistic antibody and the adverse event of fatal immune reaction. The alternative-hypothesis would be that there is some connection between the two events. If scientific data provides evidence that the alternative-hypothesis is not true, John claims that this does not mean that the null-hypothesis ought to be accepted, i.e. asserting that there is no connection between administering a super agonistic antibody and the adverse event of fatal immune reaction[19]. He suggests that investigators are accustomed to apply statistical parameters at a default setting preventing to generate false positives, i.e. data showing that there is some connection when there is no connection. By doing so, they take for granted that more false negatives are generated, i.e. data showing that there is no connection when there is in fact some connection. John concludes "that when scientific results are transferred from journals and laboratories to the context of policy and decision-making, *failure to show a link* is often treated as *proof of no link*. This is a simple epistemological error…"[20] So, Stephen John suggests that much more knowledge of probability and uncertainty of risks can be quantified. However, this requires (again) disclosure of scientific raw data of clinical experiments and this is normally not so practiced in the scientific community, which was exactly the fact that already frustrated Kenter and Cohen.

Regarding the fourth discussion about the what kind of verdict is delivered by RBA, it can be said that utilitarianism is a much disputed moral theory that has some attractive features, but faces one more cardinal problem as well. Its point of departure is primarily focused on the state of affairs that an action brings about. But,

[19] John S., Risk and Precaution, Public Health Ethics (chapter 4), ed. Angus Dawson, Cambridge University Press, 2011, p79
[20] idem

it is not clear whether RBA should adopt this focus on consequences of actions alone. There may be other important considerations as well. These considerations do not focus on the state of affairs before and after the action alone, but judge the action itself as laudable or illaudable. For example, Stephen John argues that some actions are judged because of the accompanying risks with specific moral status. He explains and makes it plausible that people value differently about avoiding harm on the one hand and relieving from suffering at the other hand. John notes that by “lay”-reasoning, members from the public attach more important moral status to avoidance of harm, rather than to relieve from suffering. He argues that this difference in moral status is caused by the fact that harming is something we must not do and suffering is something we must prevent. Accepting John’s claim means that, contrary to what was presumed in the previous section regarding the calculation sheet and applying equations literally, the weighing of risks and opportunities can only be lined up symbolically because apparently, risks weigh different from opportunities. Risks and opportunities are incommensurable in the sense that a surplus cannot be literally calculated. This problem may be partly solved by the suggested constraints of not allowing the risks of experiments to exceed comparable levels of risks of daily life standards. However, John’s argument may imply that RBA can be divided into separated analyses, a risk analysis and an opportunity analysis, and each analysis is provided with specially assigned requirements. For a risk analysis there are requirements for each individual risk; for an opportunity analysis there is the requirement of maximization of utility.

In this section I criticized the utilitarian interpretation of RBA. First, the classical version of utilitarianism encounters the sacrifice problem; more contemporary versions of utilitarianism, like pluralistic utilitarianism, encounter the problem of a higher level of complexity and are therefore less practical to apply. One feature of the version of pluralistic utilitarianism, that I suggested, is to applying additional requirements for separate risks of harm. However, it is not so evident how to use the norms or burdens for these single risks of harm in RBA. It is unclear how a daily life

standard can be applied as a limit and what justification there is for variations abroad. Second and third, it is a hard job to list all potential harms, to abandon non-realistic harms and to process scientific data regarding risks of harm in such a way that it facilitates RBA properly. Kenter and Cohen call on for the preparation of standard lists of questions leading to efficient survey of relevant potential harms[21] and Stephen John calls on for scientists to change the burden of proof when scientific results are transferred from journals and laboratories to the context of policy and decision-making. All emphasize the importance of disclosing scientific data to the scientific community. However, disclosure is a major shift for institutes and will not be implemented collectively in the short term. Fourth, it is unlikely that states of affairs alone have moral value. Actions may have intrinsic value as well. As avoidance of harm can be conceived as an action with different moral status compared to an action that relieves from suffering, risks of harm are found incommensurable with opportunities of benefits. This means that a good verdict, based on RBA, cannot be calculated as a surplus of opportunities over risks. In the next section I will indicate what can be concluded from this thesis and which recommendations can be developed.

[21] Recently (June2012), the government review board of the Netherlands (CCMO – The Hague) adopted literally Kenter and Cohen's proposed set of factors that facilitate rational risk analysis of all new substances to be administered to human beings (see http://www.ccmo-online.nl/main.asp?pid=25&sid=49&ssid=247):

a. Level of knowledge about mechanism of action
b. Previous exposure of human beings with the test product(s) and/or products with a similar biological mechanism
c. Can the primary or secondary mechanism be induced in animals and/or in ex-vivo human cell material?
d. Selectivity of the mechanism to target tissue in animals and/or human beings
e. Analysis of potential effect
f. Pharmacokinetic considerations
g. Study population
h. Interaction with other products
i. Predictability of effect
j. Can effects be managed?

6. Conclusions and the way forward

After reviewing several research protocols and after some scarce hours of witnessing the execution of RBA in review boards, I was disappointed about its results and the structure, completeness and consistency that were provided in its documents. By these observations I was undoubtedly biased to presume that more organizational effort is needed to get this extremely complex job done. Given its necessity for decision-making and its complexity in execution, two main conclusions can be put forward. First, a systematic search for risks is necessary in order to provide a useful RBA. Second, if RBA is applied according to a utilitarian interpretation, two modifications are suggested. Pain and pleasure from classical (hedonistic) utilitarianism are abandoned and replaced by Quality-Adjusted-Life-Years (QALYs), and a constraint on the maximum allowable risks is introduced in order to avoid participants from being sacrificed in experimentation. After having presented a most plausible version of a utilitarian interpretation of risk-benefit analysis, a substantial package of criticism remains. This criticism underpins the claim that risk-benefit analysis is a tool under development. If attention is given to some weak points of its procedure, risk-benefit analysis may support the balancing of potential harms and opportunities to some degree. However, it is undesirable to trust risk-benefit analysis as an indisputable calculation model commanding a decision in clinical experimentation by its own force.

Regarding the research question (*How must we understand RBA, given its necessity for decision-making and its complexity in execution?)* I conclude that it is reasonable to divide RBA into separated analyses, a risk analysis and benefit analysis, and to provide each analysis with specially assigned requirements. In response of the empirical sub-question (*How should RBA be performed according to current guidelines and how is RBA actually practiced?)* I conclude that prescriptions in clinical research guidelines lack concrete articulations of weighing of risks and

benefits. I also conclude that an account of chance of success (for benefits) is neglected. Further, it is not clear what role is ascribed to the investigator and what role is ascribed to review boards regarding RBA. I finally conclude that in 11 clinical research projects the research community somehow fails to perform a robust RBA. Apparently, RBA is a difficult task to accomplish and plausibly, review boards struggle with the topic as well. Considering the ethical sub-question (*How must we understand the utilitarian interpretation of RBA and is such an interpretation defensible?)* I conclude that it is difficult to defend utilitarian interpretations of RBA. The classical version of utilitarian interpretation of RBA yields a very simple decision model by which calculation of maximum utility and coming to a verdict seems an easy job (if scientific data are available). However, this version is not realistic because of the sacrifice problem that is accompanied by classical utilitarianism. The more contemporary version of pluralistic utilitarian interpretation of RBA yields a more complex decision model accompanied by other problems (assessment of single risks). A question can be raised whether this variant is practically applicable. A strong objection against a utilitarian interpretation of RBA, regardless of its version, is that its main parameters – risk and opportunity – are incommensurable. This makes RBA somehow lame for producing good or bad ready-made verdicts. However, if structurally applied and supported by scientific data, it is still a powerful tool for reporting the acceptability of individual risks, for the maximization of opportunities and for deciding between alternatives. So currently, RBA in clinical research is a somewhat imperfect method to work with and therefore, it must not be applied blindly. But, there is no alternative. Investigators and review boards are bound to it and must try to make use of it, at least by letting it line up their scientific convictions.

Based on the considerations in this thesis there are some recommendations that can be put forward.

First, investigators must literally present their RBA proposal during the review board sessions in which their research protocols are reviewed. This recommendation is intended to commonly improve discussion and exchange of RBA judgments including their motivations. Intensified communication yields better expertise for both investigators and review boards regarding the articulation of arguments and the understanding of the weighing process. Along this presentation of research protocols by the subsequent investigator, it must be clear what the responsibilities are of the investigator (give advice) and the review board (make judgments).

Second, it is recommended to perform separate risk-analyses and opportunity-analyses. This separation contests the suspicion that single risks might be crossed out by opportunities. Both analyses have different requirement-profiles and this separation improves clarity on the weighing-process. In this process separate judgments must be provided for risks, for opportunities and for the overall-balancing.

Third, it is recommended to apply RBA according to the ICH guideline on Quality risk management (or equivalent) and to make usage of checklists like the list provided by Kenter and Cohen. Risks should be presented in a consistent and orderly manner to boost the chance that a complete set of risk-scenarios is evaluated. An inventory of risk-scenarios is done by listing what might go wrong. If one or a few risk-scenarios are overlooked, an immense gap may exist in the RBA. So, it is arguable that standard lists with default risk-scenarios are made publicly available that can be used as starting-point to enhance the completeness and consistency of RBA. These lists are to be used and applied by all colleague researchers.

Fourth, it is recommended that review boards verify that investigators, who advise on their research protocols in the subsequent review board sessions, report the results of their experiments according to regulations of the Helsinki Declaration[22].

[22] B20: “Authors, editors and publishers all have ethical obligations with regard to the publication of the results of research. Authors have a duty to make publicly available the results of their research on human subjects and are accountable for the completeness and accuracy of their reports. They should adhere to accepted guidelines for ethical reporting. Negative and inconclusive as well as positive results should be published or otherwise made

This should be done for successful as well as for unsuccessful investigations. The reports should be easily available for other researchers in order to prevent duplication, but also to learn from failures. Especially regarding failed experiments it is important to disclose scientific data that proved such investigations unsuccessful. As indicated in section 5, this scientific data can be helpful in determining or even quantifying the risk of harm in new investigations. The verification of investigator's reporting by the review board should be carried out in a formal way. Non-compliancy should be made publicly available, at least to other review boards.

Fifth, it is recommended to make an inquiry of public opinion of acceptable single risk (chances of occurrence as well as harm levels) and to make an inquiry of public opinion of what decisions are allowed to be taken by participants and which information is appreciated by the participants. These inquiries must yield insight into the way the public appreciates a more paternalistic approach or an autonomous approach and whether the RBA itself should be disclosed. There are arguments to restrict the RBA information disclosure and to protect participants from information overload. Disclosing RBA may also conflict with the research objectives, and finally, investigators may be very detached from disclosing sensitive information because of scientific competition. However, from an autonomy perspective it could be suggested, as indicated by the Belmont Report, that all information from the RBA should be linked to the informed consent procedure. I believe the proposed inquiries provide review boards with insight how to deal with the public opinion in general and how to deal with single risks in particular.

---

publicly available. Sources of funding, institutional affiliations and conflicts of interest should be declared in the publication. Reports of research not in accordance with the principles of this Declaration should not be accepted for publication."

Literature

- Appelbaum PS, Lidz CW. The therapeutic misconception. In: Emanuel ed. The Oxford Textbook of Clinical Research Ethics. Oxford, New York: OUP 2008: 633-644.

- Beauchamp T.L., Childress J.F., Principles of Biomedical Ethics, Oxford University Press, 2009.

- Council for International Organizations of Medical Sciences (CIOMS). International ethical guidelines for biomedical research involving human subjects. Geneva: CIOMS, 2002.

- Edwards SJL, Kirchin S, Huxtable R. Research ethics committees and paternalism. J Med Ethics 2004;30:88-91.

- EJ Emanuel, D Wendler, C Grady, What makes clinical research ethical?, JAMA, 2000

- Garrard E, Dawson A. What is the role of the research ethics committee? Paternalism, inducements, and harm in research ethics. J Med Ethics 2005;31:419-423.

- Glannon W., Phase I oncology trials: why the therapeutic misconception will not go away. J Med Ethics 2006;32:252-255.

- International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use, ICH harmonised tripartite guideline, Quality risk management, Q9, step 4, 2005

- John S., Risk and precaution, Public Health Ethics (chapter 4), ed. Angus Dawson, Cambridge University Press, 2011, 67-84

- Kenter M J H, Cohen A F, Establishing risk of human experimentation with drugs: lessons from TGN1412, The Lancet, Vol 368, 2006.

- King NM., Churchill LR., Assessing and comparing potential benefits and risks of harm. The Oxford textbook of clinical research ethics. Edited by Emanuel E, Grady C, Crouch RA, Lie RA, Miller FG, Wendler D. New York: Oxford University Press; 2008:514– 26.

- Manson N. Consent and informed consent. In: Ashcroft RE, Dawson AJ et al (eds). Principles of Health Care Ethics. Second Edition. Chichester: Wiley, 2007: 297-304.

- Rid A, Emanuel EJ, Wendler D. Evaluating the risks of clinical research. JAMA 2010;304(13):1472-1479.

- Rid A., Wendler D., Risk–benefit assessment in medical research—critical review and open questions, Law, Probability and Risk 2010;9 (3-4): 151-177.

- Spriggs M., Canaries in the mines: children, risk, non-therapeutic research, and justice, Journal of Medical Ethics 2004, 30, p.180.

- Sreenivasan G. Does Informed Consent to Research Require Comprehension? Lancet 2003;362: 2016–18.

- The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report. Ethical principles and guidelines for the protection of human subjects of research, 1979.

- Timmons M., Moral Theory, An Introduction (Lanham: Rowman & Littlefield, 2002).

- Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10, Vol. 2, pp. 181-182. Washington, D.C.: U.S. Government Printing Office, 1949.

- Weijer C, Miller PB. When are research risks reasonable in relation to anticipated benefits? Nature medicine 2004;10(6): 570-573.

- Wendler D, Grady C. What should research participants understand to understand they are participants in research? Bioethics 2008;22(4): 203-208.

- Wendler D, Miller FG. Assessing research risks systematically: the net risks test. J Med Ethics 2007;33:481-486.

- World Medical Association (WMA). World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. Amendment 59th WMA General Assembly, Seoul 2008.