Setting risk thresholds in biomedical research:

Lessons from the debate about minimal risk

Annette Rid, MD

Department of Social Science, Health & Medicine

King's College London

Address for Correspondence:
Dr Annette Rid
Department of Social Science, Health & Medicine
King's College London
Strand
London WC2R 2LS
Phone: +44 207 848 7113
Fax: +44 207 848 6550
E-Mail: annette.rid@kcl.ac.uk

Keywords: minimal risk, research ethics, research ethics review, research regulation, risk-adapted regulation, risk category, risk threshold, subject protection

Word count: 9159 words

ABSTRACT

One of the fundamental ethical concerns about biomedical research is that it frequently exposes participants to risks for the benefit of others. To protect participants' rights and interests in this context, research regulations and guidelines set out a mix of substantive and procedural requirements for research involving humans. Risk thresholds play an important role in formulating both types of requirements. First, risk thresholds serve to set upper risk limits in certain types of research (e.g. pediatric research that offers the participating children no prospect of clinical benefit). Second, risk thresholds serve to demarcate risk categories that streamline risk-adapted systems of ethical oversight (e.g. expedited or no prospective ethical review of minimal risk research). But although risk thresholds play such an important role in research governance, there is a need both to better define the existing risk thresholds and to delineate new thresholds in order to develop more risk-adapted systems of research oversight. The present paper examines the existing minimal risk threshold and surrounding debates with the goal of deriving a systematic approach to setting thresholds of research risk.

(177 words)

1. INTRODUCTION

One of the fundamental ethical concerns about biomedical research is that it frequently poses risks to participants for the benefit of others. While investigational interventions (e.g. drugs or diagnostic tests) have some prospect of clinical benefit for participants, many other research procedures do not. For example, most clinical trials involve procedures that are not clinically indicated but serve to answer research questions (e.g. imaging procedures, biopsies). Moreover, entire strands of study require performing procedures purely for research purposes (e.g. phase 1 trials in healthy volunteers, natural history studies). The main benefit from these procedures is the generalizeable knowledge to be gained from the research, intended to improve our understanding of health and disease or the care of future patients.

The fact that research frequently poses risks for the benefit of others makes it important to protect the rights and interests of study participants. Essentially all research regulations and guidelines use risk thresholds to achieve this goal. Risk thresholds are used, firstly, to limit the risks to which participants may be exposed when their voluntary informed consent is not possible or feasible to obtain and the research offers them no prospect of clinical benefit. For example, many regulations and guidelines allow no more than minimal risks in "non-beneficial" research with participants who cannot give their own informed consent (e.g. children, incapacitated adults) (U.S. Department of Health and Human Services (DHHS) 1991, Emanuel et al. 2000, Council for International Organizations of Medical Sciences (CIOMS) 2002, South African Medical Research Council (SAMRC) 2002, Council of Europe (CoE) 2005, Indian Council of Medical Research (ICMR) 2006, Ad hoc group for the development of implementing guidelines for Directive 2001/20/EC (Ad hoc group) 2008, Schweizerische Akademie der Medizinischen Wissenschaften (SAMW) 2009,

Schweizerische Eidgenossenschaft, Bundesamt für Gesundheit (BAG) 2011, World Medical Association (WMA) 2013), though a "minor increase" over minimal risk is sometimes allowed in certain types of pediatric research (DHHS 1991, CIOMS 2002, Ad hoc group 2008). Similarly, many regulations and guidelines require that the risks to participants be no greater than minimal when waivers or modifications of informed consent are necessary for methodological reasons (e.g. research involving deception) or when the research could not practicably be carried out without a waiver (e.g. medical records research) (DHHS 1991, CIOMS 2002, ICMR 2006, BAG 2011, WMA 2013).

Secondly, risk thresholds are used to demarcate risk categories in order to calibrate ethical oversight to the risks that research studies pose to participants. For example, some guidelines and regulations allow expedited or no research ethics committee (REC) review for certain types of minimal risk research, while full review is required for research involving greater than minimal risks (DHHS 1991, ICMR 2006). The underlying idea is that low-risk research requires less ethical scrutiny because it is less likely to significantly compromise participants' rights and interests. "Risk-adapted" systems of ethical review – so the idea – thus allow regulators to minimize ethical oversight and the associated costs (e.g. administrative burden, potential delay or deterrent of valuable research) within the constraints of adequate subject protection (Rid 2014).

Although risk thresholds play such an important role in research governance, there are two problems with how they are currently conceptualized and implemented. The first problem regards the definition and implementation of the existing risk thresholds. Regulations and guidelines define risk thresholds (e.g. minimal risk threshold) in different ways, and this can complicate and potentially delay or stifle

multinational research projects. Moreover, empirical data suggest that the same risk thresholds are implemented differently across research studies and sites (Lenk et al. 2004, Shah et al. 2004). For example, of 188 REC chairpersons in the U.S., 23% categorized a non-beneficial allergy skin test in healthy 11-year-olds as a minimal risk procedure, 43% categorized the same procedure as a minor increase over minimal risk, and 27% categorized it as more than a minor increase over minimal risk (Shah et al. 2004). This raises concern that participants are not adequately protected in some studies, while other studies are rejected out of unwarranted concerns about research risks. A systematic approach to setting risk thresholds – which would offer a conceptually robust, empirically grounded and methodical approach to setting risk thresholds across the spectrum of biomedical research – would help to address these concerns.

The second problem with how risk thresholds are currently conceptualized regards their lack of nuance. Many regulations and guidelines only use the minimal risk threshold to stratify risks to participants; this is arguably a relatively crude way of structuring a spectrum of research risks that ranges from essentially no risks (e.g. analysis of existing, anonymous and non-reidentifiable data) to significant risks of harm (e.g. deep brain injection of a new transgene under general anesthesia). As a consequence, the current system of research regulation and oversight is often regarded as not being risk-adapted enough and commentators are proposing to delineate additional risks thresholds to address this concern (e.g. *de minimis* threshold (Rhodes et al. 2011)). These proposals are likely to be influential, given that regulators in numerous countries are currently developing more risk-adapted systems of research oversight (Rid 2014). In addition, some have urged for more guidance on risk limits in research with competent consenting participants (London 2006, Miller & Joffe

2009, Rid & Wendler 2011, Resnik 2012). A systematic approach to setting risk thresholds would greatly facilitate and harmonize the definition and implementation of new thresholds.

Against this backdrop, the present paper aims to develop a systematic approach to setting thresholds of research risk by examining the existing minimal risk threshold and the surrounding debates. Because the minimal risk threshold is widely used in international guidelines (CIOMS 2002, CoE 2005, WMA 2013), national regulations (DHHS 1991, BAG 2011), and national guidelines (SAMRC 2002, ICMR 2006, Ad hoc group 2008), it has attracted a lot of commentary.[1] It is therefore instructive to analyze the existing definitions of minimal risk and discuss their implications for setting risk thresholds in research generally. After some preliminary remarks about the concept of risk, the following discussion draws on prominent guidance and selected commentary to identify the existing strategies for setting the minimal risk threshold. Each strategy is critically discussed with a view to "lessons to be learned". The paper then combines these lessons to derive a systematic approach to setting thresholds of research risk.

Before beginning the analysis, three brief remarks about the scope of the paper are in order. As mentioned above, the existing minimal risk threshold pertains almost exclusively to the risks of research procedures and studies that offer participants no prospect of clinical benefit.[2] The following discussion therefore has the same scope. Readers with an interest in setting risk thresholds or defining risk categories in

[1] For orientation, a PubMed search with the search terms ((minimal AND risk) AND ethic*) retrieved 359 publications on 25 April 2014 (database searched since its inception). The author reviewed these publications based on their title and identified 44 papers that directly addressed the minimal risk threshold in research involving humans.

[2] U.S. regulations are a notable exception, as they define a small class of "beneficial" research interventions as minimal risk. These are interventions that they are already marketed and used within the conditions of licensure (e.g. comparative effectiveness study of two marketed interventions). See http://www.hhs.gov/ohrp/policy/expedited98.html (accessed on 25 April 2014).

"beneficial" research are asked to turn elsewhere (Rid 2014). In addition, the following discussion only includes those approaches to setting the minimal risk threshold that have potential wide applicability to setting risk thresholds across the spectrum of biomedical research. For example, according to the "moral education" standard, research risks are minimal when they are no greater than the risks to which children may be exposed for educational purposes in family life situations (Ackerman 1980). This standard is not included because it does not seem applicable beyond the context of pediatric research, while the present paper aims to derive a systematic approach to setting risk thresholds in research in general. For instance, it is difficult to see how the moral education standard could inform risk limits in research with competent adults who consent to study participation.[3] Finally, the paper inevitably takes a normative stance on thresholds of acceptable research risk by developing a particular approach to setting such thresholds. By contrast, it does not discuss what type of ethical oversight is appropriate at a given level of risk to participants.

2. RISK AND BURDEN[4]

Any discussion of risk needs to start with clarifying basic concepts and terminology, given that the word "risk" has many meanings (Hansson 2011). The following analysis assumes the standard definition of research risk that conceptualizes risk as a function of two more basic components: 1) the likelihood that a harmful event or experience will occur as a result of undergoing a research procedure, and 2)

[3] In principle, the same applies to the "scrupulous parents" standard that defines risks as minimal when informed and scrupulous parents would consent to exposing their child to these risks (Freedman et al. 1993). However, the standard assumes that scrupulous parents consider research risks as minimal when they do not exceed the risks their child faces in everyday life. Given this, relevant passages regarding the scrupulous parents standard are incorporated into the discussion of the "risks of daily life" standard that is endorsed by many regulations and guidelines (Section 3.3).

[4] This section draws considerably on two previous publications by the author (Rid & Wendler 2010, Rid & Wendler 2011). Unless otherwise noted, the statements in this section are based on these papers.

the extent to which the event or experience, should it occur, harms the individual participant. For example, an allergy skin test poses a 13- 26 per 100'000 risk of experiencing a mild systemic allergic reaction with self-limiting hay fever symptoms or urticaria requiring anti-histamine treatment. These symptoms arguably constitute a small harm, given that they are associated with moderate discomfort, short in duration, easily mitigated, and generally have no effect on the participant's ability to perform daily life activities and pursue life goals (Rid et al. 2010).

There is widespread agreement that risk judgments should reflect different types of potential harm to participants, including physical, psychological, social, and economic harms. However, existing regulations and guidelines take different views on how potential harms relate to the burden, discomfort, or inconvenience of participating in research. While most guidelines make no distinction between these terms, some guidance documents and commentators recommend treating risks and burdens separately (Levine 1986, CoE 2005, Kind 2009, Westra et al. 2011, WMA 2013). For example, the Declaration of Helsinki requires that non-beneficial research involve no more than minimal risk *and minimal burden* [emphasis added] when participants are incapable of giving informed consent (WMA 2013).

Making explicit reference to the burdens of participating in research is helpful insofar as they can be easily overlooked. Indeed, some commentators argue that being burdened or experiencing discomfort constitutes a mere inconvenience, not a harm. Moreover, because we should protect participants from excessive risks of harm, they recommend that mere inconveniences be excluded from risk judgments (Levine 1986). However, this approach is problematic because a sufficient number of burdens or inconveniences can lead to considerable harm, either as the result of additive or synergistic effects (e.g. hourly blood draws for several days can add up to a

considerable risk of harm although the risks of a single blood draw are low). Furthermore, what constitutes a burden is not clearly defined (e.g. there is no clearly defined point at which pain converts from being a burden to being a harm). Different raters may therefore exclude different burdens from their risk judgments, undermining the goal of protecting participants' rights and interests consistently across studies. It is therefore helpful to emphasize that potential harms of all magnitudes should be considered, including those that might be regarded as mere inconveniences.

However, it does not follow that research risks and burdens should be evaluated differently. Conceptually, burdens or inconveniences constitute setbacks to participants' interests just like other harms. The relevant question is whether joining a study makes a participant's life go worse, and the answer seems to be "yes" for both burdens and potential harms. For example, completing a lengthy, boring questionnaire and experiencing a systemic allergic reaction from an allergy skin test both make a participant's life go worse – but completing a questionnaire does so much less than experiencing an allergic reaction. This suggests that the difference between burdens and harms is a matter of degree. Burdens set back less important interests than harms (e.g. interest in not being bored versus interest in being physically fit), or they thwart a given interest to a lesser degree (e.g. brief and mild pain versus long and excruciating pain). The separate evaluation of research burdens and risks – understood as potential harms – should therefore be avoided.[5]

3. SETTING THE MINIMAL RISK THRESHOLD: FOUR APPROACHES

[5] For a discussion of why the difference between *certain* burdens and *potential* harms (i.e. risks) is not morally salient, see (Rid & Wendler 2011).

With these definitions and concepts in mind, we can now turn to analyzing how prominent ethical guidelines, research regulations and commentators currently set the minimal risk threshold. Four basic strategies emerge.

3.1 Procedural approach

Some regulations and guidelines set the minimal risk threshold without defining it. For example, Swiss regulations allow no more than minimal risks in non-beneficial research when participants cannot give informed consent, but they do not define what constitutes minimal risk (BAG 2011). Other guidelines equally use the minimal risk threshold without defining it (Shaddy & Denne 2010, WMA 2013). This approach effectively leaves the definition of the minimal risk threshold to those who are charged with applying it. Providing no definition thus amounts to a procedural approach to setting thresholds of research risk: a research intervention or study poses a given level of risk when an appropriately charged individual or body has made this determination.

3.1.1 Lessons to be learned

To better understand this procedural approach to setting risk thresholds, it is useful to consider the role of decision procedures for realizing right outcomes. In the present context, what is the role of procedures for ensuring that risk thresholds are set and implemented at an appropriate level? The most advanced discussions of the relationship between outcomes and procedures are found in the context of distributing scarce resources (e.g. organs for transplantation), and the different conceptions of "procedural justice" that have been put forward illuminate the relationship between

outcomes and procedures in other contexts. It is therefore helpful to draw on these conceptions for the present analysis.

There are four conceptions of procedural justice that differ depending on whether an independent criterion (or set of criteria) for the just distributive outcome exists (Rawls 1971: 83-90, Rid 2009). Two conceptions assume an independent and clearly defined criterion of justice that allows determining the just outcome in any given case; the decision procedure is only instrumental for realizing this outcome. The difference is that the procedure either works in all cases (perfect procedural justice) or in most cases (imperfect procedural justice). For example, when dividing a cake, announcing that the smallest piece will go to the person cutting it always ensures an equal division – provided the person wants as large a piece as she can get, is capable of cutting the cake equally, and so on. By contrast, following the appropriate procedures and rules of evidence in criminal trials typically ensures that the guilty are convicted, although an innocent person is sometimes sadly found culpable. The fact that both perfect and imperfect procedural justice assume clear criteria of justice quickly reveals that neither conception reflects the relationship between outcomes and procedures in the present context. When regulations or guidelines set risk thresholds without defining them, there is no independent and clearly defined criterion for limiting research risks that a perfect or imperfect procedure could realize.

The two other conceptions of procedural justice assume that independent criteria of justice are either lacking or indeterminate, and procedures therefore become more important for realizing just outcomes. When there is no independent criterion for the right outcome but an appropriate decision procedure exists, any outcome resulting from this procedure is regarded as just (pure procedural justice). For example, in a game of roulette, the distribution of money is just when it has been

determined by the roulette's spin. But charging RECs or other decision-makers with determining whether research interventions or studies fall below or above a certain risk threshold is not like playing roulette. Although vague, the thresholds offer some independent guidance simply in virtue of their existence. By contrast, reliance on pure procedural justice would imply that RECs make decisions without any guidance whatsoever. This suggests that the remaining conception of procedural justice best reflects the relationship between outcomes and procedures in the context of research regulation and oversight.

Constrained pure procedural justice assumes that an independent criterion of justice exists, but the criterion is so indeterminate that it is consistent with more than one just outcome (i.e. there are two or more equally just outcomes). When there is an appropriate procedure for choosing between these outcomes, any outcome resulting from that procedure is considered just. For example, the philosopher Norman Daniels has used this approach in his work on healthcare justice. Daniels argues that limited healthcare resources should be allocated based on Rawls' general principles of justice. But because Rawls' principles are indeterminate and reasonable people disagree about how to specify them, Daniels advocates a fair process – "accountability for reasonableness" – to choose between what he considers to be equally just allocation options (Daniels 2008, Rid 2009). Similarly, research regulations or guidelines adopt a constrained pure procedure when they set risk thresholds without defining them and charge RECs or others with their implementation. The thresholds offer independent but indeterminate guidance, and RECs then choose between equally acceptable interpretations of these thresholds in the given study context. For instance, it is widely agreed that children should only be exposed to very limited research risks when the research offers them no potential clinical benefits, but reasonable people disagree

about what this implies. That is, reasonable people who are both well-informed and motivated to find mutually acceptable ways of cooperation disagree as to where exactly a risk limit should be set.[6] By setting the minimal risk threshold in this context without defining it, regulations or guidelines reflect the idea that children should only be exposed to very limited risks. At the same time, they avoid taking a stance on a question about which reasonable people disagree.

This feature is an important advantage of setting risk thresholds through use of a constrained pure procedure. Because risk thresholds serve to limit research risks and delimit risk categories for risk-adapted ethical oversight, they are impossible to separate from contentious questions in research ethics and regulation. By setting vague thresholds, regulations and guidelines reflect the reasonable view that some risk limits need be set to effectively protect research participants while also ensuring efficient research oversight. At the same time, regulations take no definitive stance on acceptable levels of research risk or appropriate forms of ethical oversight. In this way, regulations respect that any particular risk threshold within the realm of reasonable disagreement is (at least ethically speaking) a matter of choice. This makes constrained pure procedures attractive from an ethical perspective – and, perhaps unsurprisingly, they are widely used in other contexts where reasonable disagreement is common (e.g. end-of-life decision making, priority-setting in healthcare).[7]

[6] The idea of reasonable disagreement is primarily associated with John Rawls who gives both an epistemological and a moral interpretation of such disagreement. On the epistemological interpretation, disagreement is reasonable if it is consistent with the proper use of reason in light of the available evidence. Unanimity cannot be reached because of various epistemic factors (e.g. conflicting empirical evidence, vague moral concepts, disagreement about relevant considerations or their weight, different ways of weighing competing considerations). On the moral interpretation, disagreement is reasonable when it is compatible with certain normative requirements (e.g. commitment to mutual justification or respect for citizens as free and equal) (Rawls 2005: lecture 2).

[7] Constrained pure procedures are also expedient in political or institutional contexts, as they give decision-makers the flexibility to interpret risk thresholds in different ways.

A key question is just how specific the constraints on pure procedures should be. For example, regulations or guidelines that set the minimal risk threshold without defining it give a bare minimum of guidance, as the word "minimal" is notoriously vague. Even though reasonable RECs are unlikely to judge medical records research as greater than minimal risk, or a deep brain injection with gene transfer as minimal risk, there is nothing that stops them from making these judgments when the minimal risk threshold remains undefined. Although the second example might seem especially fanciful, it is not entirely implausible on a strong relativistic interpretation of minimal risk. For example, when a brain cancer patient is certain to die within the next couple of weeks, a REC could argue that the added risk of the brain injection and gene transfer is small in relation to her high background risk of death. However, strong relativistic interpretations of risk thresholds are widely regarded as problematic because they can imply greater levels of acceptable risk for individuals or groups who are already worse off (Institute of Medicine (IOM) 2004, Kopelman 2004, Wendler 2005).[8] To exclude that RECs and others are making clearly unreasonable judgments, risk thresholds should be defined in ways that reflect the existing reasonable agreement. The above example suggests such agreement extends beyond the vague idea that research risks should be very limited in certain types of studies.

In addition, there are good practical reasons for making the existing reasonable agreement about risk thresholds as explicit as possible. Research from cognitive psychology shows that risk judgments are influenced by numerous factors

[8] As a reminder, the risk thresholds under discussion pertain to research procedures and studies that offer participants no prospect of clinical benefit and therefore impose risks purely for research purposes. This feature makes the relativistic interpretation of these thresholds so worrying. By contrast, a relativistic interpretation risk thresholds might be more defensible when research interventions have a prospect clinical benefit (e.g. greater risks or uncertainty might be acceptable when interventions have a prospect of benefit and participants suffer from a serious disease without alternative treatment options). – For clarification, the above example assumes that the gene transfer is a non-beneficial intervention. This assumption seems justified given the poor track-record of gene transfer research thus far.

(Kahneman et al. 1982, Slovic 1987, Weinstein 1989). While some factors may reflect legitimate moral intuitions about acceptable risk (e.g. benefits of the risky activity, voluntariness of those engaged, avoiding loss versus making gains), others are clearly confounders (e.g. familiarity with the activity, ease of imagining or recalling relevant examples ("availability heuristic"), framing effects). These latter factors are ethically concerning because they make risk judgments dependent on arbitrary factors. For example, the familiarity bias can imply that the same research procedure is judged to be less risky by RECs that are familiar with the procedure than by RECs that are not. In fact, the finding that risk judgments vary between RECs may at least partially be explained by this consideration (Lenk et al. 2004, Shah et al. 2004). To reduce arbitrariness and unwarranted variation in risk judgments, regulations or guidelines should define risk thresholds to the extent that reasonable agreement allows.

There are further practical reasons for specifying risk thresholds to the extent possible. Vague thresholds or constraints increase the need for robust decision procedures in terms of accountability, appeals and revision processes, stakeholder involvement, and so forth (Daniels 2008). However, robust decision procedures create administrative burden and cost that can seem disproportionate, especially when the risks to participants are limited and the research is otherwise uncontroversial. Specifying risk thresholds can reduce the need for robust procedures in order to realize acceptable outcomes. In addition, a reasonably clear definition of risk thresholds is needed to demarcate any categories of research that may not require prospective REC review. When risk thresholds are so vague that they need to be complemented with a decision process, it is difficult to designate research as exempt or excused from review. For example, it is unclear how regulations could exempt

certain minimal risk studies from REC review when they provide no definition of minimal risk.

In summary, given that risk thresholds are entangled with judgments about acceptable levels of research risk and appropriate research oversight, reasonable people are likely to disagree as to where these thresholds should be set. The best response to this situation is for regulations and guidelines to adopt a constrained pure procedure for setting risk thresholds – that is, to specify risk thresholds to the extent that reasonable agreement allows and institute appropriate procedures (e.g. ethical audits, REC review) to make decisions within these constraints. The following two sections offer ideas for how thresholds can be specified within this broad framework.

3.2 Specifying the meaning of minimal risk

Regulations, guidelines and commentators currently use two approaches to specifying the minimal risk threshold. Some give synonyms of minimal risk – such as insignificant, trivial (McCormick 1974), very low (Schweizerische Akademie der Medizinischen Wissenschaften 2009), or the least possible risk (McIntosh et al. 2000) – to further elucidate the threshold's meaning. Others revert to the two basic components of risk, likelihood and magnitude of harm, to specify what minimal risk implies. For example, the Council of Europe states that research involves minimal risks when it is to be expected that the research will result, at most, in a very slight and temporary negative health impact (CoE 2005). Conversely, some commentators propose that risks are minimal when they involve no likelihood of serious harm, permanent injury or death (Nicholson et al. 1986, Kind 2009).

3.2.1 Lessons to be learned

Of the two existing approaches to specifying risk thresholds, the use of synonyms has only limited value. Although synonyms can enrich the "feeling" for a given risk level, they offer no material guidance and therefore cannot exclude unreasonable interpretations of risk thresholds. For example, when a REC member unreasonably believes that giving a urine sample involves more than minimal risks, defining "minimal" as insignificant or very low will do little to correct his view. By contrast, referring to the two basic components of risk allows regulations or guidelines to set out more useful guidance. In particular, it becomes possible to delineate more concrete risk thresholds by specifying different combinations of likelihoods and magnitudes of harm (e.g. defining risks as minimal when they involve no likelihood of serious harm (Nicholson et al. 1986, Kind 2009)). Explicitly referring to likelihoods and magnitudes of harm also prompts raters to consider the available empirical data about the risks of research procedures, which should mitigate the influence of arbitrary factors on their risk judgments, such as the familiarity bias (Rid et al. 2010).

In order to specify risk thresholds based on likelihood and magnitude of harm, it is important to be clear about these concepts and their relationship to each another. The existing definitions of the minimal risk threshold illustrate several common confusions in this regard. As discussed in the introduction, the magnitude of harm depends on various factors that need to be considered in combination. However, regulations or guidelines frequently give excessive weight to some of these factors. In particular, many documents stipulate that permanent or irreversible harms are always serious. This is problematic because some permanent harms are small (e.g. inconspicuous scar) while some reversible harms are serious (e.g. severe pain for an extended but circumscribed period of time). While guidance should draw attention to

the factors that can influence the magnitude of harm, it should refrain from designating certain factors as overriding. Similarly, many regulations or guidelines focus exclusively on serious harms although other magnitudes of harm are also relevant. For example, a research procedure that involves a high likelihood of moderate harm arguably qualifies as greater than minimal risk, even when it is highly unlikely to result in serious harm. This shows that risk thresholds should address different magnitudes of harm and specify likelihood thresholds for each magnitude.

Furthermore, to ensure that likelihood thresholds are properly understood, some research ethicists and risk communicators recommend that they be anchored with numeric information (Nicholson et al. 1986, Rid et al. 2010, Fischhoff et al. 2011). Qualitative descriptors of likelihood thresholds (e.g. low, medium, high) are known for being interpreted in highly variable ways. To avoid unjustified variation in raters' judgment, likelihood thresholds should be anchored with numerical frequencies or frequency ranges that are representative of the given threshold (e.g. 10 or 10-2- per 100'000 for a low likelihood of harm). This would not imply that likelihood thresholds are tightly fixed at the given frequency or frequency range; this would be unreasonable given the uncertainty around the risks of many research procedures.[9] Rather, the numeric information would serve as an approximate fix point that elucidates the meaning of likelihood thresholds and orients the rater's judgment.

Finally, the exclusive focus on serious harms in some regulations and guidelines illustrates a common confusion about the relationship between the likelihood and magnitude of harm. Some guidelines stipulate that minimal risk implies no likelihood of serious harm (CoE 2005). However, few activities involve

[9] The uncertainty has many sources, including poor documentation of the risks of research procedures (creating uncertainty especially about low frequency events) and potential interactive effects between research procedures and other interventions administered as part of a trial (e.g. biopsies are more likely to cause bleeding when participants are treated with anti-coagulants).

literally no likelihood of serious harm, and this includes undergoing some research procedures that are considered to be paradigm examples of minimal risk. For example, many regulations and guidelines classify a single blood draw as minimal risk although it involves an extremely low likelihood of serious injury from fainting. This suggests that requiring minimal risk procedures to pose no risk of serious harm unreasonably excludes interventions involving acceptable risks. Of course, a literal reading of this requirement might not be intended, and "no risk of serious harm" could simply signify such a low likelihood of serious harm that it can be ignored. However, if this is the case the intended meaning should be made explicit in order to avoid unreasonable interpretations of the threshold. More generally, it is important for guidelines and regulations to convey that risk thresholds are shaped by both the likelihood and magnitude of potential harms, and that applying these thresholds therefore always requires considering both components of risk.

In summary, in order to specify risk thresholds, guidelines and regulations should delineate likelihood thresholds for different magnitudes of harm and anchor these likelihood thresholds with numeric information. For example, the minimal risk threshold might be defined as involving a very low likelihood of serious harm (<1 per 100'000), a low likelihood of moderate harm (<10 per 100'000), and up to a high likelihood of small harm (>10 per 100'000).[10] Specifying risk thresholds in this way would arguably reflect areas of reasonable agreement (e.g. minimal risks can involve a very low likelihood of serious harm) while leaving room for diverging views among reasonable people (e.g. concerning the question whether an extremely low likelihood

[10] Magnitudes of harm and likelihood thresholds could usefully be anchored with existing categorizations (e.g. research harm scale of the "systematic evaluation of research risks" method (Rid et al. 2010), CIOMS categories of frequency for adverse drug reactions (CIOMS 1999).

of serious harm can offset a slightly elevated likelihood of moderate harm).[11] The following section explores how regulations and guidelines might anchor likelihood thresholds for different magnitudes of harm by comparing the risks of research to the risks of other activities.

3.3 Invoking risk comparisons

Numerous regulations and guidelines define the minimal risk threshold in comparison to the risks of activities other than research. The three most prominent comparator activities are routine clinical examinations and tests (DHHS 1991, CIOMS 2002, Ad hoc group 2008), daily life activities (DHHS 1991, Freedman et al 1993, SAMRC 2002, ICMR 2006, Ad hoc group 2008), and charitable participation (Wendler 2005). For example, the CIOMS guidelines define research risks as minimal if they are no more likely and not greater than the risks attached to routine medical or psychological examinations (CIOMS 2002). U.S. regulations stipulate that risks are minimal when the probability and magnitude of harm or discomfort anticipated in the research is not greater than those ordinarily encountered in daily life (DHHS 1991). Moreover, some commentators have argued that risks qualify as minimal when they do not exceed the risks of charitable activities deemed acceptable for these individuals in daily life (Wendler 2005).[12]

3.3.1 Lessons to be learned

[11] More generally, risk thresholds raise intricate questions about the permissibility of balancing likelihoods and magnitudes of harm and limits on aggregate risks. It seems unlikely that reasonable people would agree about these issues. Moreover, even if they did, incorporating such agreement in regulatory or ethical guidance would almost certainly exceed the level of specificity with which regulations can operate.

[12] The routine examinations standard and the risks of daily life standard are either used alone (e.g. CIOMS 2002) or in combination (e.g. DHHS 1991, Ad hoc group 2008). In fact, routine examinations can reasonably be seen as a particular daily life activity.

The previous section concluded that risk thresholds should be specified by delineating likelihood thresholds for different magnitudes of research harm and anchoring these likelihood thresholds with numeric information. But how should likelihood thresholds be anchored? Existing guidance and commentary suggests that analogical reasoning is helpful in this regard.

Analogies are commonly used in moral deliberation in order to orient and focus our judgment about new or contested questions. If we have a considered judgment about the right course of action in a particular case, and that case is relevantly similar to the case under discussion, we can infer through analogical reasoning that both cases should likely be judged in the same way.[13] In the present context, if we have a considered judgment that the risks of a given activity are acceptable, and the activity is relevantly similar to research, we can infer that similar risks are likely acceptable in the research context. Furthermore, if we know the risks associated with the comparator activity, we can use this information to anchor likelihood thresholds for different magnitudes of research harm. For example, we may have a considered judgment that the risks of participating in a charity soccer game are acceptable for children, and we may further believe that this activity is relevantly similar to research because it poses risks primarily for the benefit of others. In that case, we can evaluate the available evidence on the risks of playing soccer in children and determine the approximate likelihood with which they will experience small, moderate and serious harm from engaging in this activity. These likelihood estimates

[13] Considered judgments are provisional fixed points in our reasoning that are characterized by a high degree of confidence in judgment (i.e. we cannot readily imagine that a considered judgment would be shaken).

can then be used to anchor likelihood thresholds for research harms of comparable magnitudes.[14]

Reasoning by analogy has important benefits (Sunstein 1993, Lamond 2006). We are often more confident in our judgments about particular cases than we are about general theories that attempt to account for these judgments. Many therefore regard the use of analogies as a more productive way to approach new or contested ethical questions. This may especially apply in the present context, given that the ethics of risk is a relatively new area of inquiry and a general theory of acceptable risk still remains to be developed (MacLean 2012). Furthermore, analogies are useful heuristic devices for deepening and sharpening our judgment, while exposing us to a wider variety of considerations than the particular case before us. For example, we may have the considered judgment that the risks of riding a car – a common daily life activity – are acceptable because these risks are normally assumed voluntarily, reasonably reduced (e.g. safety standards for cars and roads), offset by the individual and social benefits of transportation (e.g. supplying basic needs, enabling social cohesion or intellectual exchange), and distributed relatively evenly within the population. We can then scrutinize the salience of these factors in the research context, repeat the process for other potential comparator activities, and thereby test and refine our intuitions about acceptable research risks.

In addition, analogies form an important basis for our judgment about particular cases and are central for developing general principles and theories (Sunstein 1993). For instance, comparing research risks to the risks of driving and other activities may gradually generate general principles for the acceptability of

[14] Some commentators justify the reference to non-research risks on other grounds. For example, Freedman and colleagues have suggested that research risks are minimal when they displace the risks that participants would have faced in daily life absent the research (i.e. the research does not make them worse off in risk terms) (Freedman et al. 1993). The present paper does not discuss the merits of such alternative justifications for risk "comparisons".

research risks. But even if general principles are impossible to generate (e.g. because they cannot account for the diversity or plurality of relevant factors), analogical reasoning promotes consistency in how we judge risks across different areas of life or requires justification for why we might evaluate research risks differently. Overall this advances our reasoning about research risks and helps to establish the boundaries of reasonable agreement about risk thresholds in research.

For such analogical reasoning to be useful, it is essential to further specify comparator risks. As many commentators have pointed out, merely identifying broad categories of comparator risks as a benchmark (e.g. the risks of daily life activities, routine examinations or charitable participation in the case of the minimal risk threshold) offers insufficient guidance and can have problematic implications (Kopelman 2000, Resnik 2005, Wendler 2005). Such broad categories comprise very different levels of risk depending on who engages in the given comparator activity, and under what conditions. For example, "routine" examinations range from undergoing a non-invasive physical exam as part of an annual check-up (posing very low risks) to undergoing a liver biopsy as a liver cancer patient (involving significant risks). "Daily life" activities range from sitting on the couch to rock-climbing or working as a high sea fisherman. And "charitable participation" can involve administering donations from a desk or providing humanitarian assistance in a war zone.

Given that risk thresholds serve to demarcate upper limits of research risk and categories for risk-adapted oversight, it is important to specify comparators that involve risks on the upper boundary of the "normal" range. For example, a blood draw might fall in the upper boundary for routine examinations, but not a physical exam (normal but not risky) or a liver biopsy (risky but not normal). Riding a car

could be suitable for daily life activities, but not watching TV or rock climbing. And playing soccer for a charity fundraiser might be appropriate for charitable participation, but not administering donations or assisting in war zones. The obvious question then is what defines the upper boundary of the "normal" range. Many commentators advocate a statistical interpretation of the "normal" and index comparators to riskier activities that average, healthy, normal individuals undertake (IOM 2004, Kopelman 2004, Wendler 2005). Furthermore, some propose age-relative comparator activities because the "normal" varies significantly with age, and this variation reflects morally salient characteristics of individuals in different age groups (e.g. ability to consent) (Wendler 2009).

The statistical approach has the advantage of excluding relativistic interpretations of the "normal" that reasonable people would likely reject, given that these interpretations imply greater levels of acceptable research risk in individuals who are already worse off. For example, the statistical approach would exclude judging a non-beneficial liver biopsy to be a low-risk research procedure in liver cancer patients simply because these patients routinely undergo liver biopsies.[15] However, the approach faces the problem that it is unclear why the statistical average should have normative force.[16] Furthermore, defining comparator risks as those that average, healthy individuals face in different age groups only specifies comparator risks to some extent. In particular, the definition only excludes certain relativistic interpretations of the "normal" – namely those within societies at a given point in time (Freedman et al 1993). It does not exclude relativistic interpretations between societies because average, healthy individuals face different levels of risk in different

[15] The example assumes that the liver biopsy would be performed purely for research purposes, which makes a relativistic interpretation of comparator risks so concerning. See also footnote 8.

[16] One possibility is that the life of the average, healthy, normal individual has normative force because it is thought to go reasonably well. However, to the author's best knowledge this argument has not been made and would require further scrutiny than is possible here.

societies. Similarly, it does not exclude relativistic interpretations over time because the risks of activities pursued by average individuals evolve due to technological and societal developments (e.g. driving a car is less risky today than it was 50 years ago). While these issues require further exploration (e.g. Snyder et al. 2011), specifying comparator risks as those that average, healthy, normal individuals in different age groups face in riskier activities excludes relativistic interpretations of comparator risks within societies that reasonable people would likely reject. The definition therefore takes an important step towards specifying risk thresholds to the extent that reasonable agreement allows.

Despite the advantages of risk comparisons and the possibility of reasonably specifying comparator risks, it is important to recognize the two fundamental limitations of analogical reasoning in the present context. The first limitation regards the difficulty of establishing that the comparison between the risks of research participation and other activities is in fact useful. Every activity is in some ways similar to, and in some ways different from, participating in research. Any comparison will therefore reveal both similarities and dissimilarities; however, too many dissimilarities or the wrong kinds of dissimilarity make risk comparisons useless or even lead to problematic conclusions. The important question therefore is what level and type of similarity between research participation and comparator activities we need in order to consolidate our judgment about risk thresholds in research. Answering this question requires identifying the morally salient features of research participation and evaluating to what extent potential comparator activities share these features. But the difficulty is that both the morally salient features of research and the needed level of similarity between research and comparator activities remain to be clarified.

It is widely agreed that biomedical research generates generalizable knowledge primarily to improve the health of future people, and that this orientation towards others is morally salient (Emanuel et al. 2000, CIOMS 2002, Kopelman 2004, Wendler 2005, London 2006, Miller & Joffe 2009). Charitable participation and other activities aimed at helping others (e.g. volunteer emergency assistance or donating blood) have the same orientation, and for this reason they are attractive comparator activities for setting thresholds of research risk (Wendler 2005, London 2006, Miller & Joffe 2009). By contrast, routine examinations and daily life activities (e.g. riding a car, playing sports) are less compelling as comparators because individuals engage in them mostly to promote their own interests. However, some dissimilarities between participating in research and engaging in charitable activities, donating blood, or providing emergency assistance remain. For example, while participating in research may or may not benefit future people, the three latter activities are likely to benefit identifiable individuals who exist today. Further analysis is needed to evaluate to what extent these and other dissimilarities compromise the analogy between research and other activities oriented towards addressing the needs of others.[17] In the meantime, specifying comparator risks as those involved in other-directed activities establishes similarity with a feature of research that is widely regarded to be morally salient.[18]

The second limitation of analogical reasoning regards the difficulty of establishing that the risks of comparator activities are themselves acceptable. The

[17] Commentators have suggested that other features of research participation might influence risk comparisons (e.g. study participants assume risks in a principle-agent relationship (London 2006), participants passively submit to the direction of a professional who is charged with a protective responsibility towards them (Miller & Truog 2009)).

[18] Even though some groups (e.g. infants, frail elderly) cannot actively engage in other-directed activities, they may appropriately be involved in them in a passive capacity. For example, it is appropriate for a father to take his infant on a car ride while he is taking an injured neighbor to the hospital. This suggests that the risks of other-directed activities remain a useful comparator even when the population in question cannot actively engage in these activities.

comparison between participating in research and another activity is useful only when we believe the risks of that activity to be acceptable. Yet the difficulty is that judgments about acceptable risk in comparator activities can be subject to reasonable disagreement just like in research. For instance, although most people consider the risks of (responsible) driving to be acceptable, some vehemently reject this idea. In addition, even considered judgments about the risks of comparator activities may be driven by arbitrary factors or the demands of rationality, not morality. For example, we may consider the risks of driving for a charity fundraiser acceptable simply because driving is a familiar activity and such an important part of our life that it would be irrational to contemplate the associated risks before every car ride. This suggests that the acceptability of risks in comparator activities can be difficult to ascertain.

One response to this problem is to try and formulate independent normative standards for the risks of comparator activities. For example, it has recently been argued that the daily life risks of children are acceptable when they do not compromise a child's ability to achieve sufficiently high degrees of the substantive goods of childhood (Binik 2014). Trying to formulate such independent normative standards is important, as judgments about particular cases should count as "considered" only when they survive the encounter with principles and theories of various levels of generality. However, it is not clear that we can achieve reasonable agreement about independent standards of acceptable risk in comparator activities. In fact, recent work on ethics and risk has revealed the numerous difficulties of developing such standards (MacLean 2012). Furthermore, even if reasonable people can agree on independent standards of acceptable risk, it is questionable that these standards are determinate enough to guide judgments about acceptable risk in

particular activities (e.g. driving as part of a charitable fundraiser). This suggests that analogical reasoning will always have a role to play. The hope and expectation is that some judgments about the risks of comparator activities will be more considered than our intuitions about acceptable risk in research, and analogical reasoning allows us to draw on these more considered judgments. It is important to recognize, however, that such reasoning only offers some degree of moral certainty about risk thresholds in research.[19]

In summary, risk comparisons are useful for specifying thresholds of research risk to the extent that reasonable agreement allows, and they offer a way of anchoring likelihood thresholds for different magnitudes of research harm. At present, comparator risks are reasonably specified as the risks that average, healthy, normal individuals in different age groups face in riskier but still acceptable activities that are directed at benefiting others. For example, playing a game of soccer in a charity fundraiser might be an appropriate comparator activity for research involving children and mowing the lawn for an infirm neighbor could be appropriate for research with adolescents. Upper risk limits in research with competent consenting adults might be anchored in comparison to the risks of providing emergency assistance or donating a kidney. And providing humanitarian assistance could be an appropriate comparator activity for setting risk thresholds in research in emergency situations (e.g. pandemic) that involves competent consenting adults.

3.4 Giving Examples

The use of examples is a common feature of definitions of the minimal risk threshold. Currently the examples for minimal risk procedures fall into two groups.

[19] Additional uncertainty may result from the fact that the risks of specific comparator activities (e.g. playing soccer) are often poorly documented.

The first group includes procedures that pose essentially no additional risks when compared to routine clinical practice (e.g. obtaining small quantities of blood or tissue during clinically indicated procedures (McIntosh et al. 2000, Ad hoc group 2008)). The second group includes non-invasive procedures (e.g. measurement of weight (U.S. Office for Human Research Protections (OHRP) 1998, U.S. Secretary's Advisory Committee on Human Research Protections (SACHRP) 2005), collection of urine (McIntosh et al. 2000, SACHRP 2005, Ad hoc group 2008), ultrasound (OHRP 1998, Ad hoc group 2008), electroencephalogram (OHRP 1998, Ad hoc 2008)). Only a few guidelines list heel or finger pricks and blood draws as examples, although they typically do not specify the acceptable number of procedures or amount of blood to be taken (SACHRP 2005, Ad hoc group 2008). To the author's best knowledge, other invasive procedures are currently not included.

3.4.1 Lessons to be learned

Examples of research procedures that pose a certain level of risk are extremely valuable. They allow regulations and guidelines to show how risk thresholds apply to different procedures and can help to establish default judgments about the risks of common research procedures. For example, a guidance document for implementing the 2001 EU Clinical Trials Directive lists electroencephalogram and electrocardiogram among the procedures that typically qualify as minimal risk (Ad hoc group 2008).

To realize the benefits of giving examples, it is important for regulations and guidelines to take several steps. Guidance documents should clarify whether the examples are primarily illustrative or intended to establish default judgments about the risks of the given procedures. For instance, the document for implementing the

Clinical Trials Directive states clearly that its risk classifications can be used as default judgments but need to be interpreted in the given study context (Ad hoc group 2008). Regulations and guidelines should ideally give default classifications of the risks of all common research procedures (e.g. various imaging procedures, biopsies) in different age groups. Investigators and RECs could then interpret these defaults in the given circumstances. For example, an allergy skin test is arguably a minimal risk procedure in average, healthy adults, but it can involve greater than minimal risks when research participants have a history of anaphylactic shock. This illustrates that default classifications always require interpretation. Nevertheless, they have the important advantage of orienting investigators' and RECs' judgment and thereby streamlining the ethical review process (Rid et al. 2010).

If it is not feasible to set defaults for all common procedures in different age groups, regulations and guidelines should try to provide at least two or three default classifications of procedures that demarcate the upper boundary of the given risk threshold. For example, the current lists of minimal risk procedures primarily include interventions that clearly qualify as minimal risk (e.g. height measurements) while offering no examples of interventions on the upper boundary of the minimal risk threshold (e.g. allergy skin testing). The lists therefore offer only limited guidance for making minimal risk determinations in borderline cases. This problem is exacerbated by the fact that current guidelines do not indicate how the listed procedures were judged to be minimal risk. It is therefore difficult for investigators and RECs to evaluate the validity of these judgments and understand when they are justified in deviating from default risk judgments, or how they should judge the risks of procedures that are not included in the list (e.g. rare or novel research procedures). Therefore, when using examples to set default risk classifications, regulations and

guidelines should aim to cover the most common procedures and explain how they arrived at their judgments, based on the available evidence and the methodology used.

4. SETTING RISK THRESHOLDS: A SYSTEMATIC APPROACH

The existing definitions of the minimal risk threshold in research regulations or guidelines and the surrounding debate offer a rich resource for thinking about risk thresholds in research in general. The final section of this paper combines the "lessons learned" from the minimal risk threshold with the goal of delineating a general approach to setting risk thresholds in research.

Risk thresholds primarily serve to set upper risk limits (e.g. in pediatric research) and demarcate risk categories that streamline risk-adapted systems of ethical oversight (e.g. expedited REC review for low-risk research, full REC review for other research). Both functions are entangled with judgments about acceptable levels of research risk and appropriate forms research oversight, and reasonable people disagree about many of these issues. At the same time, risk judgments in research (and elsewhere) are susceptible to various cognitive biases that can cause arbitrariness and unwarranted variation of these judgments. A key challenge therefore is to set risk thresholds in a way that respects reasonable disagreement, while providing adequate guidance to minimize unreasonable judgments about research risks.

The "lessons learned" from the minimal risk threshold suggests that this challenge is best met when regulations and guidelines adopt a "constrained pure procedural" approach to setting risk thresholds. This implies that regulations specify risk thresholds to the extent that reasonable agreement allows and institute procedures (e.g. ethical audits, REC review) for evaluating risks within these constraints. To specify risk thresholds, guidelines and regulations should refer to the two basic

components of research risk, likelihood and magnitude of research harm. Furthermore, they should distinguish different magnitudes of harm (e.g. small, moderate, serious) and set approximate likelihood thresholds for each magnitude of harm. Likelihood thresholds are best anchored with numeric information (e.g. < 1 per 100'000) that is derived from risk comparisons, thereby drawing on considered judgments about acceptable risk in activities that are relevantly similar to research. For example, if the risks of participating in a charity soccer game are considered acceptable, and this activity is thought to be relevantly similar to research, data on the risks of playing soccer can be used to anchor likelihood thresholds for different magnitudes of research harm. At present, comparator risks are best specified as the risks that average, healthy, normal individuals in different age groups face in riskier but still acceptable activities that are directed at benefiting others (e.g. playing soccer for charitable purposes in children, providing emergency assistance in adults). Finally, guidelines and regulations should establish default judgments about the risks of common research procedures that investigators and RECs then evaluate in light of the particular study under consideration. To enhance the value of such defaults, an explanation should be given regarding how the defaults were set.

Specifying risk thresholds in this way arguably respects reasonable disagreement about acceptable research risks and appropriate research oversight, while reflecting areas of reasonable agreement. When used to refine existing risk thresholds and specify new thresholds, this approach should make an important contribution to developing regulatory frameworks that both effectively protect study participants and enable efficient ethical oversight. At the same, it is important to recognize that certain fundamental moral uncertainties about thresholds of research risk will remain.

Acknowledgements

Many thanks to Frank Miller and David Wendler for helpful discussion and comments. The research leading to these results has received funding from the Swiss Academy of Medical Sciences (Käthe-Zingg-Schwichtenberg Fonds 13/10) and the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement n° 301816. A previous version of this paper was awarded the 2012 Mark S. Ehrenreich Prize in Healthcare Ethics Research (Runner-up) by the International Association of Bioethics and the University of California, Pacific Center for Health Policy and Ethics.

6. REFERENCES

Ackerman, T. F. (1980). Moral duties of parents and nontherapeutic clinical research procedures involving children. Bioethics Quarterly, 2(2), 94-111

Ad hoc group for the development of implementing guidelines for Directive 2001/20/EC relating to good clinical practice in the conduct of clinical trials on medicinal products for human use (2008). Ethical considerations for clinical trials on medicinal products conducted with the paediatric population. European Journal of Health Law, 15(2), 223-250

Binik, A. (2014). On the minimal risk threshold in research with children. American Journal of Bioethics (in press)

Council for International Organizations of Medical Sciences (CIOMS) (1999). Reporting adverse drug reactions: definitions of terms and criteria for their use. (Geneva: CIOMS)

Council for International Organizations of Medical Sciences (CIOMS) (2002). International Ethical Guidelines for Biomedical Research Involving Human Subjects. (Geneva: CIOMS)

Council of Europe (CoE) (2005). Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research. (Strasbourg: Council of Europe)

Daniels, N. (2008). Just Health: Meeting Health Needs Fairly. (Cambridge, New York: Cambridge University Press)

Emanuel, E. J., Wendler, D., & Grady, C (2000). What makes clinical research ethical? Journal of the American Medical Association, 283(20), 2701-2711

Fischhoff, B., Brewer, N. & Downs, J. (Eds.) (2011). Communicating Risks and Benefits: An Evidence-Based User's Guide. (Silver Spring, MD: Food and Drug Administration)

Freedman, B., Fuks, A. & Weijer, C. (1993). In loco parentis. Minimal risk as an ethical threshold for research upon children. Hastings Center Report, 23(2), 13-19

Hansson, S. O. (2011). Risk. (In E.N Zalta (Ed.), The Stanford Encyclopedia of Philosophy (SEP) (Spring 2014 Edition). Stanford: SEP.)

Indian Council of Medical Research (ICMR) (2006). Ethical Guidelines For Biomedical Research On Human Participants. (New Dehli, India: ICMR)

Institute of Medicine (IOM) (2004). Ethical Conduct of Clinical Research Involving Children. (Washington, D.C.: The National Academies Press)

Kahneman, D., Slovic, P. & Tversky, A. (1982). Judgment under uncertainty: heuristics and biases. (Cambridge, New York: Cambridge University Press)

Kind, C. (2009). Evaluation of risk in research with children - it's time to clear the misconceptions. Bioethica Forum, 2(2), 74-79

Kopelman, L. M. (2000). Children as research subjects: a dilemma. Journal of Medicine and Philosophy, 25(6), 745-764

Kopelman, L. M. (2004). Minimal risk as an international ethical standard in research. Journal of Medicine and Philosophy, 29(3), 351-378

Lamond, G. (2006). Precedent and analogy in legal reasoning. (In E.N Zalta (Ed.), The Stanford Encyclopedia of Philosophy (SEP) (Spring 2014 Edition). Stanford: SEP)

Lenk, C., Radenbach, K., Dahl, M. & Wiesemann, C. (2004). Non-therapeutic research with minors: how do chairpersons of German research ethics committees decide? Journal of Medical Ethics, 30(1), 85-87

Levine, R. J. (1986). Ethics and Regulation of Clinical Research. (Baltimore: Urban & Schwarzenberg)

London, A. J. (2006). Reasonable risks in clinical research: a critique and a proposal for the Integrative Approach. Statistics in Medicine, 25(17), 2869-2885

MacLean, D. (2012). Ethics and risk. (In S. Roeser, R. Hillerbrand, P. Sandin & M. Peterson (Eds.), Handbook of Risk Theory (pp. 791-804). Heidelberg, New York: Springer.)

McCormick, R. A. (1974). Proxy consent in the experimentation situation. Perspectives in Biology and Medicine, 18(1), 2-20

McIntosh, N., Bates, P., Brykczynska, G., Dunstan, G., Goldman, A., Harvey, D., Larcher, V., McCrae, D., McKinnon, A., Patton, M., Saunders, J. & Shelley, P. (2000). Guidelines for the ethical conduct of medical research involving children.

Royal College of Paediatrics, Child Health: Ethics Advisory Committee. Archives of Disease in Childhood, 82(2), 177-182

Miller, F. G. & Joffe, S (2009). Limits to research risks. Journal of Medical Ethics, 35(7), 445-449.

Nicholson, R. H. & Institute of Medical Ethics (Great Britain) (1986). Medical research with children: ethics, law, and practice. The report of an Institute of Medical Ethics working group on the ethics of clinical research investigations on children. (Oxford, New York: Oxford University Press)

Rawls, J. (1971). Theory of Justice. (Harvard: Harvard University Press)

Rawls, J. (2005). Political Liberalism (expanded edition). (New York: Columbia University Press)

Resnik, D. B. (2005). Eliminating the daily life risks standard from the definition of minimal risk. Journal of Medical Ethics, 31(1), 35-38

Resnik, D. B. (2012). Limits on risks for healthy volunteers in biomedical research. Theoretical Medicine and Bioethics, 33(2), 137-149

Rhodes, R., Azzouni, J., Baumrin, S. B., Benkov, K., Blaser, M. J., Brenner, B., Dauben, J. W., Earle, W. J., Frank, L., Gligorov, N., Goldfarb, J., Hirschhorn, K., Hirschhorn, R., Holzman, I., Indyk, D., Jabs, E. W., Lackey, D. P., Moros, D. A., Philpott, S., Rhodes, M. E., Richardson, L. D., Sacks, H. S., Schwab, A., Sperling, R., Trusko, B. & Zweig, A. (2011). De minimis risk: a proposal for a new category of research risk. American Journal of Bioehtics, 11(11), 1-7

Rid, A. (2009). Justice and procedure: how does "accountability for reasonableness" result in fair limit-setting decisions? Journal of Medical Ethics, 35(1), 12-16

Rid, A. (2014). How should we regulate risk in biomedical research? An ethical analysis of recent policy proposals and initiatives. Health Policy (in press). DOI 10.1016/j.healthpol.2014.04.011

Rid, A., Emanuel, E. J. & Wendler, D. (2010). Evaluating the risks of clinical research. Journal of the American Medical Association, 304(13), 1472-1479

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

Rid, A. & Wendler, D. (2011). A Framework for Risk-Benefit Evaluations in Biomedical Research. Kennedy Institute of Ethics Journal, 21(2), 141-179

Schweizerische Akademie der Medizinischen Wissenschaften (SAMW) (Ed.) (2009). Forschung mit Menschen: Ein Leitfaden für die Praxis. (Basel: Schwabe AG)

Schweizerische Eidgenossenschaft, Eidgenössisches Departement des Inneren, Bundesamt für Gesundheit (BAG) (2011). Bundesgesetz über die Forschung am Menschen (Humanforschungsgesetz, HFG) vom 30. September 2011. (Bern: BAG)

Shaddy, R. E. & Denne, S. C. (2010). Clinical report--guidelines for the ethical conduct of studies to evaluate drugs in pediatric populations. Pediatrics, 125(4), 850-860

Shah, S., Whittle, A., Wilfond, B., Gensler, G. & Wendler, D. (2004). How do institutional review boards apply the federal risk and benefit standards for pediatric research? Journal of the American Medical Association, 291(4), 476-482

Slovic, P. (1987). Perception of risk. Science, 236 (4799), 280-285

Snyder, J., Miller, C.L. & Gray, G. (2011). Relative versus absolute standards for everyday risk in adolescent HIV prevention trials: Expanding the debate. American Journal of Bioethics, 11(6), 5-13

South African Medical Research Council (SAMRC) (2002). Guidelines on Ethics for Medical Research: General Principles, including research on children, vulnerable groups, international collaboration and epidemiology. (Tygerberg, South Africa)

Sunstein, C. (1993). On analogical reasoning. Harvard Law Review, 106, 741-791

U.S. Department of Health and Human Services (DHHS) (1991). U.S. Code of Federal Regulations, 45 CFR 46; revised 1991

U.S. Department of Health and Human Services, Office for Human Research Protections (OHRP) (1998). Categories of Research That May Be Reviewed by the Institutional Review Board (IRB) through an Expedited Review Procedure. Federal Register, 63(216), 60364-60367

U.S. Department of Health and Human Services, Secretary's Advisory Committee on Human Research Protections (SACHRP) (2005). Appendix B. Internal document to provide clarification to the recommendations relating to 45 CFR 46.404, 46.405 and 46.406. Retrieved 23 May, 2014, from http://www.hhs.gov/ohrp/sachrp/sachrpltrtohhssecapdb.html

Weinstein, N. (1989). Optimistic biases about personal risks. Science, 246, 1232-1233

Wendler, D. (2005). Protecting subjects who cannot give consent: toward a better standard for "minimal" risks. Hastings Center Report, 35(5), 37-43

Wendler, D. (2009). Minimal risk in pediatric research as a function of age. Archives of Pediatrics & Adolescent Medicine, 163(2), 115-118

Westra, A. E., Wit, J. M., Sukhai, R. N. & de Beaufort, I. D. (2011). How best to define the concept of minimal risk. Journal of Pediatrics, 159(3), 496-500

World Medical Association (WMA) (2013). WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. (Ferney-Voltaire, France: WMA

Penultimate draft – please email for final version: Rid A: Setting risk thresholds in biomedical research: lessons from the debate about minimal risk. *Monash Bioethics Review (forthcoming)*

Table. Setting risk thresholds in research: a systematic approach.

Step	Justification	Example
1) Recognize overall goal of setting risk thresholds: specify risk thresholds to the extent that reasonable agreement allows and institute appropriate procedures to evaluate risks within these constraints	• Reasonable people agree that some risk thresholds need to be set to effectively protect research participants and enable efficient ethical oversight, and they agree about the broad contours of risk thresholds. However, they disagree as to where precisely risk thresholds should be set • Judgments about research risks can be influenced by arbitrary factors (e.g. framing effects) and therefore risk being unreasonable	• Merely setting the minimal risk threshold without defining it does not exclude clearly unreasonable judgments (e.g. liver biopsy is a minimal risk procedure). Therefore specify the threshold to the extent possible
2) Specify risk thresholds using the two basic components of research risk: distinguish different magnitudes of research harm and specify likelihood thresholds for each magnitude	• Reminds raters that all risks are a function of both the likelihood and magnitude of harm • Encourages raters to consider the available evidence on the risks of research procedures, thereby reducing the influence of arbitrary factors on their judgment	• The minimal risk threshold might be defined as involving a very low likelihood of serious harm, a low likelihood of moderate harm, and up to a high likelihood of small harm
3) Anchor likelihood thresholds for different magnitudes of research harm with numeric information: draw on data about the risks of appropriate	• Numeric information reduces unjustified variation in how likelihood thresholds are understood • Drawing on data about the risks of appropriate	• Participating in a charity soccer game might be an appropriate comparator activity for research involving children. If the data show that children face a

	comparison to other activities allows drawing on considered judgments about acceptable risk in other areas of life. It also promotes consistency in risk judgments across different areas of life	for moderate harm in pediatric research
4) Ensure that the choice of comparator activities is indeed appropriate: comparator risks should reflect risks that average, healthy, normal individuals (i) in different age groups (ii) face in riskier (iii) but still acceptable activities (iv) that are directed at benefiting others (v)	• (i) avoids exposing individuals who are already worse off (e.g. seriously ill) to greater risks • (ii) acknowledges age as a proxy for morally relevant features (e.g. ability to consent) • (iii) ensures useful guidance by defining the upper boundaries of risk thresholds • (iv) ensures that judgments about acceptable risk in comparator activities are considered • (v) ensures that comparator activities are relevantly similar to research	• The following comparator activities might be acceptable: ∗ Charitable soccer in children ∗ Charitable lawn mowing in adolescents ∗ Emergency assistance or live kidney donation in adults ∗ Humanitarian assistance in adults during crisis (e.g. pandemic)
5) Develop default judgments about the risks of common research procedures and explain how these defaults were set	• Streamlines judgments about the risks of common research procedures • Enables raters to understand how risk thresholds are applied, decide when to deviate from defaults, and apply the risk threshold to rare or novel procedures	• As a default, allergy skin testing in healthy adults is a minimal risk procedure (given the available data and use of a systematic method for comparing the associated risks with the risks of appropriate charitable activities)