Check for updates

Foundations of site selection procedures for deep geological repositories: an argument-based model to explain how site rejection decisions can be justified by inaccurate operationalizations and assessments of long-term protection

Martin Navarro*

Department F, Federal Office for the Safety of Nuclear Waste Management (BASE), Cologne, Germany

Site selection procedures for deep geological repositories are driven by the rejection of candidate sites whose degree of long-term protection is insufficient or less sufficient. If long-term protection is defined in relation to future exposures, it has to be operationalized, that is, translated into measurable indicators, such as *dose* or *degree of containment*, which, again, have to be evaluated by safety assessments. Site selection procedures, therefore, depend on the quality with which long-term protection is operationalized and assessed. Although it is widely acknowledged that operationalizations and assessments of long-term protection are inherently inaccurate, little attention has been paid to the question whether these inaccuracies prevent site selection procedures from improving long-term protection. Still, there is no theory of site selection that could specify the conditions under which site selection procedures are rational with regard to the target of long-term protection. To contribute to such a theory, a conceptual model is presented that explores how site rejection decisions can be justified by inaccurate operationalizations and safety assessments. The model rests on the assumption that site rejections are justified by logical arguments. By explicating what is needed to support the arguments, the model displays the complex structure of the justification, which, amongst others, rests on the quality of operationalization, safety assessment and system understanding. The presented argument-based approach is novel in the context of site selection. However, it is not meant as an alternative to multi-criteria decision-making, but as a necessary complement to understand the potential and limitations of safety-related decision criteria. The presented model identifies which types of errors are tolerable in the context of site selection and it explains why error tolerance is lowest for safety comparisons. The model points out that the frequently used assessment strategy of conservatism is not suitable for rejecting sites for reasons

of insufficient or lower safety. It also shows that consensual requirements for the conditions under which long-term protection is achieved may be powerful tools for site selection.

1 Introduction

1.1 Aiming for long-term protection

In many countries, deep geological repositories are the preferred option for the disposal of high-level radioactive waste (IAEA, 2011). Site selection procedures for deep geological repositories are driven by the rejection of candidate sites whose degree of long-term protection is insufficient or less sufficient. In Germany, for example, site selection is a step-wise process in which candidate regions or sites are rejected in accordance with the German Site Selection Act (BT, 2023). The overarching goal is to find a 'site with best possible safety' (Liebscher et al., 2020; Hoyer et al., 2021) and for the purpose of this article, we will assume that the term 'site' does not only refer to a geographical or geological domain but also the implemented safety concept and repository design, which are crucial to the safety of a site.

In Germany, the most general meaning of safety is the long-term protection of people and the environment for 1 million years (BT, 2023). Generally speaking, the term 'protection' may have different meanings: For example, it may refer to future exposures to ionizing radiation or to the quality of the measures that are taken to prevent future exposures. These two interpretations differ in meaning since the taken measures need not succeed in limiting exposures. While the IAEA's explanation of 'protection' (IAEA, 2022) does not clearly prefer one of these interpretations, the ICRP is more specific. It defines protection quantities as dose quantities 'that allow quantification of the extent of exposure of the human body to ionising radiation' (ICRP, 2007). This definition clearly connects protection to exposures, which is reflected by the common practice of calculating dose estimates for future populations. The particular problem of connecting long-term protection to future exposures is that the latter cannot be measured directly. The present study will nevertheless follow this interpretation of long-term protection and investigate the problem.

1.2 The need for a safety assessment perspective

Site selection for geological repositories is a decision-making problem that involves multiple decision criteria. Multi-criteria decision analysis is, therefore, often recognised as an appropriate approach (Gutberlet, 2015; Madeira et al., 2016; Anelli et al., 2025) although practical implementation may be challenging (Department for Communities and Local Government, 2009). Since site selection procedures try to improve or reach a sufficient degree of long-term protection, at least some decision criteria must be related to long-term protection. These will be referred to as 'safety criteria'.

The rationality of the site selection process strongly depends on whether the chosen safety criteria are suitable for improving long-term protection. Assessing this suitability is not an objective of multi-criteria decision analysis. Instead, it requires a safety assessment perspective. While, in the context of multi-criteria decision-making it may be sufficient to justify safety criteria by expert judgement, national regulations or international guidelines (Merkhofer and Keeney, 1987; Petraš, 1997; Taji et al., 2005; Schwenk-Ferrero and Andrianov, 2017; Bilgilioğlu, 2022; NAGRA, 2024), the safety assessment perspective needs to clarify the scientific reasoning behind the choice of safety criteria. This is what the presented study tried accomplish and the reason why it did not adopt a multi-criteria decision-making approach.

1.3 The problem of connecting indicator values with future exposures

From the perspective of safety assessment, the justification of safety criteria depends on the quality with which long-term protection is operationalized, that is, translated into measurable safety, performance or safety function indicators (OECD-NEA, 2012). The suitability of indicators, therefore, does not only depend on properties like communicability, measurability and practicability (Heiermann and Olszok, 2024) but also on the relation between the chosen set of indicators and long-term protection.

Establishing such a relation proves difficult if long-term protection is related to future exposures. With regard to the well-established safety indicator *dose*, for instance, the International Commission on Radiological Protection (ICRP) pointed out that 'current judgements about the relationship between dose and detriment may not be valid for future populations' (Valentin, 1997). This raises doubts as to whether dose indicators can measure future exposures in principle, and this is usually expressed by the statement that 'dose calculations are no predictions' (Ewing and Grambow, 2025).

Attempts at solving the problem have been made by arguing that dose indicators are no measures of safety, thereby implying that establishing a close relation between dose and future exposures might not be necessary. In accordance with this argumentation, the OECD-NEA considered the estimated dose to be 'only an indicator of safety and not a measure of safety' (OECD-NEA, 2002) so that it can be used 'for testing against regulatory and design targets' (ibid.). However, this does not relieve the indicators and associated targets from connecting with future exposures in order to be meaningful. Consequently, it remains necessary to specify this connection instead of leaving it vague. This is also true if indicators are prescribed and, thus, justified by regulations. The reason is that if limiting future exposures is indeed a regulatory goal, regulators

should be able to explain how the prescribed indicators are connected to future exposures.

The most promising candidate for such a connection is a causal relation. In the context of site characterization and safety assessment, indictors are, therefore, often regarded as suitable if the measured physical quantities exert a considerable effect on future exposures. This holds for many geological indicators (Turner et al., 2023; Müller et al., 2024), such as *hydraulic conductivity*, *porosity*, *age of fluid inclusions* (Mallants et al., 2024) or *repository-induced effects on the long-term stability of the geological and engineered barriers* (Papafotiou et al., 2022), as well as for indicators that evaluate the repository's overall behavior, such as *degree of containment* or *dose* (BT, 2023).

Yet there are reasons not to be content with causal relations alone: As will be shown, they do not guarantee that the indicators values are arguably connected to future exposures. In other words, they do not guarantee that the operationalization and assessments of long-term protection are sufficiently *accurate*. We will use the term 'accuracy' here as a general term to refer to the suitability for making reliable statements about the level of long-term protection or its boundaries. The term is introduced for the sake of brevity and to provide a term that does not only refer to the errors of measurement and safety assessment, but also to those of operationalization. We will try to further specify the meaning of 'accuracy' in this article.

1.4 Sources of inaccuracy

Before we pursue the question, which accuracy is needed in the context of site selection, we will first ascertain how acute and ubiquitous the problem of inaccuracy is. As the following examples show, inaccuracies are connected to the choice, evaluation and aggregation[1] of indicators, and in many cases, they should be difficult to quantify.

Example 1 (indicator choice): Performance and safety function indicators, such as *degree of containment* and *age of fluid inclusions*, measure the fulfilment of repository functions (because performance depends on function fulfilment). However, fulfilment of functions is not equivalent to the avoidance of future exposures: Even a non-functioning repository that does not ensure containment might sufficiently limit future exposures if radionuclides are effectively dispersed in the repository's geological overburden or in the biosphere. Consequently, performance and safety function indicators are inaccurate operationalizations of future exposures (or of their absence).

Example 2 (indicator evaluation): Some indicators, such as *dose* and *degree of containment*, must be evaluated for possible evolutions of the repository system. Technically, this is achieved by assessing scenarios. Scenarios are often called *potential* or *possible* evolutions of a repository system (OECD-NEA, 2012). However, if there is a misperception of the system's real possibilities, some scenarios may be impossible. The assessment of impossible evolutions will not

---

1 In this article, 'evaluation' denotes the process of assigning values to indicators and 'aggregation' the process of weighting and combining indicators.

assess the long-term protection that is provided by the real system but subjective ignorance. Unfortunately, the inaccuracy of such an assessment cannot be quantified, as misperceptions of system possibilities may go unnoticed. Whether this problem is likely to occur depends on the nature of scenario uncertainties. Some authors consider scenario uncertainties to be mostly aleatoric (OECD-NEA, 2012; Tosoni et al., 2018; Kuhlman et al., 2024), in which case most scenarios would be real possibilities of the system. An assessment of impossible evolutions would then indeed be unlikely. However, there are compelling reasons to believe that many uncertainty contributions to scenario uncertainties are epistemic, that is, caused by a lack of knowledge of something that could be known in principle:

1. Scenarios often refer to existing heterogeneous system properties, such as heterogeneous permeability fields or properties of joint or fault networks. Even if these properties are generated by random processes, the uncertainty about them is not aleatoric, as it may seem, but epistemic: It is caused by the limited perception of properties that already exists.

2. Aleatoric uncertainty is caused by repository processes that respond sensitively to small state changes, which often applies to threshold-dominated processes like fracture generation. However, many repository processes, such as contaminant diffusion through a clay host rock or heat flow through a salt host rock, are not sensitive enough to produce such a behavior.

3. A repository system cannot be dominated by aleatoric uncertainties if claims of robustness are true. A shaft seal, for example, could never be called robust if its failure were a matter of chance. Repository systems and safety functions are usually called robust if they are insensitive (OECD-NEA, 2012; BT, 2023), which makes them more predictable. This, in fact, is the opposite of random behavior. A dominance of epistemic uncertainties should therefore be characteristic of robust systems.

On this account, epistemic uncertainties should be a major constituent of scenario uncertainty. Consequently, every comprehensive set of scenarios should contain impossible evolutions. Indicators that have to be evaluated for comprehensive scenario sets, such as *dose* or *degree of containment*, therefore cannot be accurate.

Example 3 (indicator aggregation): If performance and safety function indicators, such *porosity* or *conductivity*, are used, they must be weighted and combined to arrive at a statement about future exposures. The indicator's individual weights must reflect their relative physical effect on future exposures. This effect, however, is difficult to determine due to the complexity and uncertainty of repository processes. For example, it should be difficult to ascertain the physical importance of the indicator *porosity* relative to other (not necessarily independent) indicators. Aggregation rules, therefore, can only be rough approximations of physical relationships. Consequently, they provide inaccurate operationalizations of long-term protection.

The examples given show how difficult it is to establish an arguable connection between indicator values and future exposures

even for indicators that are established in the expert community like *material flux*, *permeability*, *mechanical stability* and *dose* (Heiermann and Olszok, 2024). This difficulty is not only caused by the uncertainties and errors that are connected to the estimation of indicator values. It is also caused by the possibility that the chosen indicators and their further processing might not be particularly suited for providing statements about future exposures. Consequently, not only the assessment but also the operationalization of future exposures can be inaccurate in the above-mentioned sense.

We now turn to the question of how accurate operationalizations and assessments of long-term protection need to be in the context of site selection. Are they accurate enough to steer a site selection procedure towards the site with the highest level of long-term protection? It is known from safety cases and other safety demonstrations (IAEA, 2012; OECD-NEA, 2012; OECD-NEA [Organisation for Economic Co-operation and Development - Nuclear Energy Agency, 2013]) that safety assessments need not be accurate predictions of long-term protection, in particular if they aim at implementing a conservative assessment strategy (Finsterle and Lanyon, 2022). However, this need not be true for site selection procedures because site rejections do not necessarily aim at demonstrating safety. Sites can also be rejected because they are not safe enough, which does not call for demonstrations of safety but for demonstrations of insufficient safety. The question of how accurate operationalizations and assessments of long-term protection need to be, therefore, should require an answer that is specific to the context of site selection. This answer requires an in-depth understanding of how site rejection decisions are justified with regard to long-term protection. Unfortunately, a theory of site selection that could improve this understanding is still not available.

1.5 Towards a theory of site selection

To contribute to the development of such a theory, a conceptual justification model is presented in this article. The model takes an integrated view on site rejection decisions, safety operationalizations and safety assessment. By doing so, it describes possible ways of justifying site rejections with regard to long-term protection and identifies acceptable types of errors. Practical conclusions are draw with regard to site selection and safety assessment.

The model construction was based on the notion that argumentation theory (Kelly and Hutchins, 2021) should play a more prominent role in justifying site rejection decisions. The model was, therefore, constructed around a set of logical arguments that may be used to justify site rejections. Starting with these arguments, additional conceptual elements were introduced that were required for their support.

With its focus on argumentation, the presented approach to decision-making in site selection procedures is novel and deviates from the common approach of multi-criteria decision-making. However, it is not meant as an alternative to multi-criteria decision-making. Rather, it should be understood as a necessary complement that allows to understand the potential and limitations of safety criteria that are used in multi-criteria decision-making. Enhancing this understanding is still necessary to prevent site

selection procedures from improving long-term protection only nominally.

The empirical basis of the presented model is the common practice of long-term safety assessment (OECD-NEA, 2012). As a representation of how site selection decisions can possibly be justified by safety assessments in practice, the model can be tested against empirical evidence and, thus, be considered a scientific model.

The model's purpose is not to provide a step-by-step guide to justifying site rejection decisions with regard to long-term protection. Instead, its main purpose is to consistently explain the nature of this justification and unfold its complex structure. However, since scientific theories and models are inductively underdetermined by empirical evidence (Okasha, 2003), the model cannot claim to provide the only possible explanation of that justification. Furthermore, it must omit some details of justification, as will be pointed out in the discussion. For this reason, the model should be regarded as a first step towards clarifying safety-related justifications of site rejection decisions.

Chapter 2 describes the justification model. It will explain the model from the perspective of safety assessment and not from that of science theory (which would be worthwhile, but was not intended for this study). Chapter 3 will highlight important practical conclusions for site selection and long-term safety assessment. Amongst others, it will point out possible ways of accelerating site selection procedures and clarify the conditions under which sites with different host rocks and safety concepts can be compared and ranked according to their level of protection, which is a specific challenge of the German site selection procedure.

2 The justification model

2.1 Basic structure

From a general point of view, the model was constructed by defining site rejection requirements and identifying the elements that are needed to satisfy these requirements. The novel aspect of the approach is that the model does not refer to regulatory or design requirements, but to the requirement that there should be sound logical arguments for rejecting a site with regard to long-term protection.

The model's main structure becomes visible if we consider how site rejection decisions relate to empirical evidence. Site rejection decisions do not draw directly on empirical evidence, but on statements about long-term protection. For example, we might reject a site because it does not seem to provide a sufficient degree of long-term protection. Long-term protection is a quantity that cannot be measured directly. Consequently, it must be operationalized, amongst others, by means of indicators. These indicators are evaluated by safety assessments that, eventually, draw on empirical evidence.

Basing site rejection decisions on empirical evidence, therefore, involves three steps, which are described by the model via the concepts *rejection argument*, *operational definition of the level of long-term protection* and *safety test* (see Figure 1). Auxiliary concepts are introduced to support these main concepts. For example, the auxiliary concept *safety dimension* is used to construct the concept
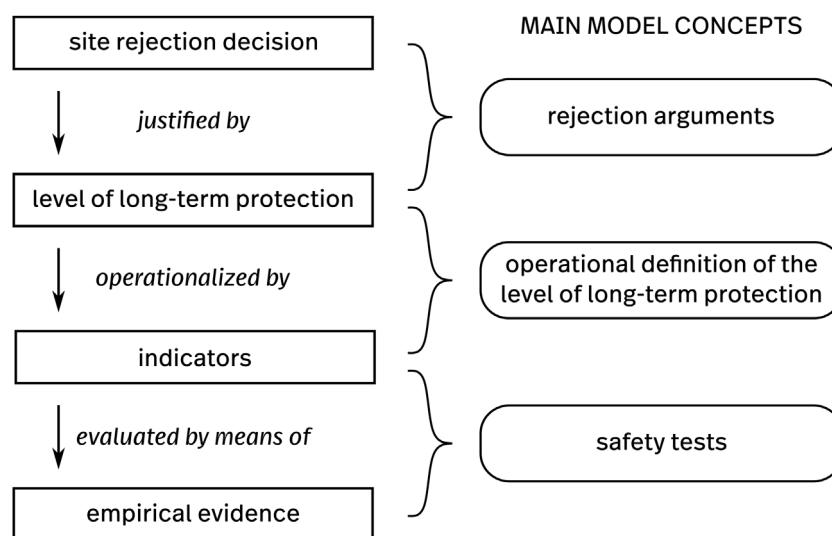
FIGURE 1
Main model concepts used to connect site rejection decisions with empirical evidence.

rejection argument. We will therefore start by introducing this auxiliary concept.

2.2 Dimensions of long-term safety

Safety is often related to protection (IAEA, 2022). In the context of site selection, safety can be described as a multidimensional property. The reason is that decision-makers may define independent site selection targets for different safety aspects, such as long-term safety, operational safety or the safety of retrieval, as it is the case in Germany. Site selection targets may differ with regulatory context and may be different for repository construction, operation and the post-closure phase.

The presented model will focus on long-term safety and assumes it to have the following safety dimensions.

1. The level of long-term protection. In Germany, this includes the protection of the environment (BT, 2023). The model is constructed on the assumption that long-term protection is a matter of future exposures. According to this assumption, the level of long-term protection is a latent (i.e., not observable) variable because future exposures cannot be measured today. (Note that the alternative assumption that long-term protection refers to the quality of the measures that are taken to prevent future exposures can also be covered by the model, as will be pointed out in the discussion.)
2. The conditions under which long-term protection is achieved. This safety dimension is, for example, addressed by the principle of containment and isolation (IAEA, 2012). The principle prescribes that long-term protection should be achieved under the condition that radionuclides are isolated and contained (and not diluted and dispersed). If sites do not implement this principle, they would not usually be accepted, even if the level of protection were sufficient.

3. The certainty of statements on the level of protection. Whether this is a safety dimension is debatable since the protection that is provided by a system does not depend on the epistemic uncertainty about that protection. However, uncertainties bear the possibility of safety deficits and are thus, in a way, related to safety. For this reason, uncertainties have been regarded as suitable criteria for site selection by Fischer-Appelt and Baltes (2010).

The model will try to connect safety assessments with statements about long-term protection. This is difficult if the terms 'safety' and 'protection' differ in meaning. It was, therefore, necessary to postulate a close connection between safety and protection. According to IAEA (2022), both terms refer to radiation risks or exposures. The term 'safety' however seems to have a broader range of application. For example, safety indicators are usually not called 'protection indicators'. For this reason, we will postulate that protection is a particular form of safety, with safety being the more general term. On this basis, it will be possible in principle to derive statements about long-term protection from statements about safety.

We will consider safety to be a continuous quantity. This opposes the view of Fischer-Appelt and Baltes (2010), who consider safety as equivalent to the achievement of protection targets. According to that view, there are no continuous levels of safety, but only the two states *safe* and *unsafe*. This dichotomous interpretation of safety, however, does not square with site selection procedures that try to maximize safety levels, even for safe sites. It also contradicts the common notion that safe repositories should become safer if barriers are added or improved.

2.3 Rejection arguments

We will now introduce the concept *rejection argument*, which is a key concept of the justification model. It is based on the notion that

it should be possible to make the logical arguments explicit that justify site rejections from a safety assessment perspective, for example, because a site cannot afford the necessary level of long-term protection.

Rejection arguments will be constructed in relation to the following safety-related site selection targets, which are based on the above-mentioned safety dimensions (numbers in brackets refer to a safety dimension):

- Ensuring a sufficient level of long-term protection (1).
- Ensuring that a sufficient level of long-term protection can be demonstrated (1, 3).
- Maximizing the level of long-term protection (1).
- Ensuring that long-term protection is achieved under acceptable conditions (2).

These targets are primary site selection targets, which may be achieved with the help of secondary targets like, for instance, the increase of scientific knowledge. The model does not claim to cover all primary site selection targets that are possible.

Site rejection decisions shall be called *justified* if the rejection is based on a rejection argument that is sufficiently sound. A rejection argument shall be called *sufficiently sound* if and only if its conclusion follows logically from its premises and if the premises are either true or sufficiently probable. It is the main intention of the presented model to identify what is needed to render rejection arguments sufficiently sound and additional concepts will be introduced to this end.

2.3.1 Rationale 1: the level of long-term protection is insufficient

Sites whose level of long-term protection is insufficient can be rejected. The rejection argument reads as follows and will be used to explain the general structure of the rejection arguments.

Premise 1: The rejection of a site is justified if its level of long-term protection is insufficient. *(general rationale)*

Premise 2: Safety test T_{nc}, which is called a *non-compliance test*, yields a positive result for site S. *(safety assessment result)*

Premise 3: If safety test T_{nc} yields a positive result for site S, its level of long-term protection is insufficient. *(operationalizing premise)*

Conclusion: The rejection of site S is justified.

The first premise holds the general rationale for site rejection. The premise is true if decision-makers accept the material implication[2]. Whether the level of long-term protection is insufficient, is determined by the following two premises and does not affect the truth value of premise 1.

The second premise claims that a safety test has yielded a positive result. A *safety test* shall be a binary classifier, which yields either a positive or negative result. These results are derived from safety assessments. Normally, safety assessments do not provide binary results. The rejection decision, however, must be based on a criterion that provides a binary result (meaning 'reject' or 'do not reject'). It is the task of the safety test to implement this criterion.

The meaning of the test result is determined by the third premise. This premise claims that the level of long-term protection is insufficient and, for this reason, the safety test is called a *non-compliance test*. The term was chosen under the assumption that non-compliance with regulation is translated to 'insufficiently safe', which need not be true in every case. The test name, however, is not relevant to the rejection argument.

The third premise infers a statement on the level of protection from a test result. The inference requires some additional information, which will be identified later. Premises of this kind shall be called *operationalizing premises* because they require an operationalization of long-term protection.

Premise 3 is uncertain. For example, there might be uncertainty as to whether the assessment tools and workflows are flawless. A more general source of uncertainty is that assessment models are underdetermined by the given body of knowledge and that assessment models are justified by auxiliary hypotheses whose truth cannot be established (Oreskes et al., 1994). The truth of operationalizing premises, therefore, can only be claimed with some degree of certainty. This degree of certainty can be expressed as a degree of belief or epistemic probability (Chuaqui, 1991) that the premise is true. It is the objective of the justification model to identify the factors that influence the epistemic probability of premise 3.

Let us now characterize the non-compliance test of the above argument. A positive test result will lead to a site rejection, whereas a negative result will not (see Table 1). For example, a non-compliance test might be testing the presence of active vulcanism. A positive test result will cause a site rejection, whereas a negative test result will not lead to any decision, which accounts for the fact that reliable safety statements cannot be inferred from a lack of active vulcanism without further information. Apparently, negative test results do not reduce the number of candidate sites. This may be problematic from the perspective of efficiency, but it is not problematic from the viewpoint of safety-related justification.

Since negative test results do not prompt any decision, they are allowed to be erroneous. This tolerance towards erroneous negative test results is important. It allows non-compliance tests to overestimate safety in order to make the safety statement that is derived from a positive test result more reliable, that is, to increase the epistemic probability of the operationalizing premise. This strategy of overestimating safety shall be called *anticonservatism*. It increases the test's positive predictive value and lowers its negative predictive value.[3]

Anticonservatism is not a common assessment approach. In the face of uncertainties, safety assessment often underestimate safety, which is usually called 'conservative'. (Note that this notion of

---

TABLE 1 Consequences of positive and negative test results if the operationalizing premise is sufficiently probable (otherwise, no decision is made).

Rejection argument	Test names	Test requirement	If test result positive	If test result negative
1	non-compliance test	anticonservative	reject site	no decision
2	compliance test	conservative	reject site	no decision
3	suboptimality test	anticonservative	reject site	no decision
3	ranking test (A < B?)	correct sign of safety difference	reject site A	reject site B
4	concept violation test	consensual requirements	reject site	no decision

conservatism is slightly broader than that of Vigfusson et al., 2007). Anticonservatism is defined here as the contrasting approach that aims at overestimating safety. If an anticonservative assessment states that a site is unsafe, we can be sure that it is because safety was overestimated. How anticonservatism should be implemented depends on the assessment approach (see chapter 3.2) and is, therefore, not part of the model description.

We will assume that it is possible in principle to identify whether an assessment model is conservative or anticonservative by investigating how the safety effect of the chosen model assumptions differs from the safety effect of more probable model assumptions.

2.3.2 Rationale 2: it cannot be demonstrated that the level of long-term protection is sufficient

It is pointless to select a site if it is sure that its safety cannot be demonstrated in the ensuing licensing procedure within a reasonable time and with reasonable use of resources. Decision-makers might, therefore, want to reject such sites.

To construct a rejection argument, we will first introduce an auxiliary licensing argument. The latter aims at demonstrating safety and, therefore, requires a safety test that checks for compliance with regulations.

Premise 1: The licensing of a site is justified if its level of long-term protection is sufficient.

Premise 2: Safety test T_c, which is called a *compliance test*, yields a positive result for site S.

Premise 3: If safety test T_c yields a positive result for site S, its level of long-term protection is sufficient. (*operationalizing premise*)

Conclusion: The licensing of site S is justified.

The operationalizing premise refers to positive test results only. So, like non-compliance tests, compliance tests require a high positive predictive value and tolerate false negatives. According to the operationalizing premise, positive compliance tests express a sufficient level of safety. In the face of uncertainties, this requires an *underestimation* of safety, i.e., a conservative assessment strategy.

We now construct the rejection argument:

Premise 1: The rejection of a site is justified if its licensing cannot be justified within a reasonable time and with reasonable use of resources.

Premise 2: No positive compliance test can be achieved for site S in a licensing procedure within a reasonable time and with reasonable use of resources.

Premise 3: If premise 2 is the case, the licensing of site S cannot be justified.

Conclusion: The rejection of site S is justified.

The argument acknowledges that resources for safety demonstrations are limited. Note that an impossibility to demonstrate safety during the site selection procedure does not imply that safety cannot be demonstrated in the ensuing licensing procedure, where resources can be focused on a single site.

2.3.3 Rationale 3: the level of long-term protection is not optimal

The following rejection argument refers to differences in safety and allows for two different types of safety comparison: *Suboptimality* and *ranking tests*.

Premise 1: The rejection of a site is justified if another site exists whose level of long-term protection is higher.

Premise 2: Safety test T_{no}, which is either called a *suboptimality test* or a *ranking test*, yields a positive result for site S.

Premise 3: If safety test T_{no} yields a positive result for site S, then another site exists whose level of long-term protection is higher. (*operationalizing premise*)

Conclusion: The rejection of site S is justified.

Suboptimality tests assess whether a considered site is not optimal without specifying alternative sites. For example, decision-makers might believe that, if the seismicity of a site exceeds a critical level, it should be possible to find safer sites.

Suboptimality tests take the form of exclusion criteria that refer to a single dominant indicator like, for instance, *seismic activity*, *groundwater age* or *thickness of the natural barrier*. Consequently, they do not differ from non-compliance tests in form, but only in the type of safety statement, which is a statement of *relative* safety.

In their formal simplicity, exclusion criteria are rough estimates and, for this reason, suboptimality tests must strongly overestimate safety. Suboptimality tests, therefore, must be *anticonservative* (we will investigate in chapter 3.2 what an anticonservative exclusion criterion is). If an anticonservative suboptimality tests yields the positive result that there should be safer sites than the one under observation, the statement will be reliable due to the overestimation of safety. However, for the same reason, a negative test result will not be reliable. In other words, suboptimality tests achieve a high positive predictive value at the price of a low negative predictive value.

Ranking tests shall be thorough, site-specific safety assessments that are used to compare and rank two specific sites according to their level of protection. We will assume that decision-makers aim for a complete ranking of all candidate sites, which they can accomplish by pairwise comparisons. When comparing two sites, A and B, the questions "Is A safer than B?" and "Is B safer than A?" should both be answered correctly. For this reason, ranking tests must also have a high negative predictive value, although this cannot be derived directly from the rejection argument. In other words, ranking tests need to assess the signs of safety differences correctly. It is important to note that this assessment does not depend on system properties or assessment errors that are shared by the compared sites and, thus, do not affect safety differences.

Please note that definition of a ranking tests does not necessarily imply that the test implements a total safety order on the set of candidate sites and that it is suitable for supporting the operationalizing premise. Which qualities ranking tests need to have will be described in chapter 2.8.

The question whether a certain site is the best candidate in a certain region (without specifying alternative sites inside that region) is not directly covered by suboptimality and ranking tests. However, the question can be answered by ranking tests if alternative sites are specified within the region. For this reason, no separate rejection argument will be introduced for this particular question.

2.3.4 Rationale 4: long-term protection cannot be achieved under acceptable conditions

Safety concepts that achieve safety mainly through dilution and dispersion of radionuclides are usually not accepted, even if they would ensure sufficient long-term protection. Hence, there may be unconditioned conceptual requirements regarding the safety dimension *conditions under which safety is achieved*. Consensual requirements of this kind provide comparatively strong arguments for site rejection because they relate to safety.

We will allow conceptual requirements to apply to all characteristics of the repository system and its environment, including system-specific uncertainties, since they all shape the conditions under which safety is achieved.

If the violation of a conceptual requirement is considered unsafe, the site can be rejected because of insufficient safety (chapter 2.3.1). If the requirement is unconditioned, so that it even holds for safe sites, the following argument can be used.

Premise 1: The rejection of a site is justified if the conditions under which long-term protection could be achieved are not acceptable.

Premise 2: Safety test T_{cv}, which is called a *concept violation test*, yields a positive result for site S.

Premise 3: If safety test T_{cv} yields a positive result for site S, the conditions under which long-term protection could be achieved are not acceptable.

Conclusion: The rejection of site S is justified.

Premise 3 is an operationalizing premise only if the concept violation tests use indicators that cannot be measured directly.

Note that concept violation tests are just one way of expressing conceptual preferences. Every cross-conceptual evaluation of safety levels attaches weights to concept-specific indicators, which reflect conceptual preferences. However, concept violation tests communicate conceptual preferences more transparently.

2.4 Operational definition of the level of long-term protection

The main challenge of achieving sound rejection arguments is to establish a sufficiently high probability of uncertain premises. The material implications of these premises draw an inference about the level of protection from a safety test result. Amongst others, this inference requires that the level of protection has been accurately operationalized. Before specifying what an accurate operationalization is, we first introduce the concept *operational definition of the level of long-term protection*, which prescribes how the level of long-term protection should be measured. The prescription is not only normative but also descriptive insofar as it should reflect the decision-maker's and regulator's understanding of safety.

An operational definition of the level of long-term protection shall consist of the following elements, some of which are optional. They allow for the practice of complementary indicators or indicator sets (OECD-NEA, 2012). An *indicator set* shall consist of indicators that require an *aggregation rule* to calculate a safety level. We will assume that indicator sets contain more than one indicator. Multiple indicator sets may be used to provide complementary assessments of safety levels.

1. *Indicators, possibly grouped into complementary sets of indicators to compensate for specific shortcomings of other indicators or indicator sets, respectively* (mandatory). For example, a dose indicator might be complemented by a containment indicator: While the latter is more loosely connected to future protection, it is less affected by uncertain processes in the biosphere and geological overburden. Some indicators may be able to indicate a safety level.

2. *An aggregation rule for each complementary set of indicators that combines the indicators into a complementary safety level indicator* (mandatory for indicator sets). For example, a set of safety function indicators might be used for deriving a safety level.

3. *An aggregation rule that combines complementary safety level indicators into an overall safety level* (mandatory if there are complementary safety level indicators). Independent safety levels might, for example, be derived from a dose indicator and from a set of safety function indicators. Since these levels can differ, they must be combined by an aggregation rule in order to provide an unambiguous safety level. The aggregation rule may include the instruction that certain complementary safety levels are neglected.

4. *Methods for evaluating indicators* (optional). These could, for example, specify scenario classes for the evaluation of a dose indicator.

5. *A list of mandatory safety factors that must be assessed, either by indicators or by the method of indicator evaluation* (optional). For example, decision-makers or regulators could demand the

assessment of certain safety factors that could act as safety reserves.

6. *Criteria for sufficient and insufficient safety formulated with regard to the specified indicators and evaluation methods* (mandatory). These might, for example, be dose limits.

We will need to define the conditions under which the operational definition of the level of long-term protection is accurate. In order to do so, we must first introduce the two auxiliary concepts *operationalizing postulate* and *system understanding*.

2.5 Operationalizing postulates

The difficulty of establishing a causal relation between dose indicators and long-term protection was already pointed out. With regard to the dose indicator, a possible solution is indicated by the ICRP's basic principle 'that individuals and populations in the future should be afforded at least the same level of protection from actions taken today as is the current generation' (Valentin, 1998). Indeed, a dose indicator could be justified under the postulate that the relationship between dose and detriment remains constant. Very likely, this postulate is not true. Yet it may be acceptable if stakeholders believe that any deviation between the postulate and reality remains small or if the postulate's error appears to be acceptable with regard to the alternative of not measuring long-term protection at all.

Accepted (but not necessarily true) postulates, which are used to establish a justifying connection between indicators and the level of long-term protection, shall be called *operationalizing postulates*. It is to be expected that operationalizing postulates are indicator-specific. For example, performance and safety function indicators require the additional postulate that function and long-term protection are equivalent.

Operationalizing postulates provide the missing link between indicators and long-term protection. By doing so, they help in exchanging negative characterizations of indicators ('dose calculations are no predictions of long-term protections') by positive characterizations ('dose calculations are predictions of the level of long-term protection or of its boundaries under the assumption that the relationship between dose and detriment remains constant'). Note that the negative characterization 'dose indicators are no predictions of long-term protection' is problematic for two reasons. Firstly, the statement remains unclear if it is not clarified what it means that something is a prediction of long-term protection (in particular, if long-term protection relates to future exposures). Secondly, the justification of dose indicators rests on the assumption that the estimated dose tells us something about the level of long-term protection or its boundaries. In this sense, dose indicators do provide predictions. Operationalizing postulates help to specify the boundary conditions of these predictions.

2.6 System understanding

Establishing causal relations between indicators and long-term protection requires an understanding of physical causes and effects and, thus, system understanding. The acceptance of containment indicators, for instance, rests on the understanding that containment will cause protection. This also holds for the contra-factual conditions of impossible evolutions.

We will define *system understanding* as a system of accepted propositions about direct or indirect physical causes of long-term protection. This includes empirical knowledge, expert judgement and propositions of accepted theories and models.

We now need to define the conditions under which the system understanding is adequate. Such a definition cannot claim consistency. The reason is that, due to the complexity of repository systems, the system understanding is composed of individual knowledge contributions, which might contradict each other (Ewing and Grambow, 2025). Moreover, theories and models may contradict empirical evidence due to idealizations and other simplifications. For this reason, we will not claim consistency, but only *sufficient justification*. This justification depends on the acceptability of inconsistencies, inaccuracies and uncertainties, which was not investigated in this study. We will nevertheless assume that decision-makers are able to assess whether a system understanding is sufficiently justified.

On this account, the *adequacy of the system understanding* shall be as the epistemic probability that the particular part of the system understanding that is used for the support of rejection arguments is sufficiently justified, which has to account for given empirical evidence and the epistemic probability that important empirical evidence is still missing.

2.7 Accuracy of the operational definition of the level of long-term protection

The *accuracy of the operational definition of the level of long-term protection* can now be defined as the epistemic probability that the following conditions are satisfied:

1. The operationalizing postulates are acceptable and the system understanding is sufficiently adequate.
2. The indicators, rules, methods and mandatory safety factors used to operationalize the level of protection are sufficiently justified with regard to the system understanding and operationalizing postulates. (Again, justification does not require consistency since operationalizing postulates, the system understanding, conservative and anticonservative assumptions, as well as idealizations and simplifications of models and theories, may introduce tolerable inconsistencies).
3. The mandatory safety factors and their physical relevance are sufficiently captured by the specified indicators, aggregation rules and evaluation methods.
4. With regard to the given safety understanding, the mandatory safety factors and evaluated indicators capture all causes of safety that need to be considered for the given rejection argument.

2.8 Accuracy of safety tests

The *accuracy of safety tests* shall be the epistemic probability that the safety test does not render the operationalizing premise false by

itself. A safety test shall be *sufficiently accurate* if it is sufficiently valid, objective and reliable.

Roughly speaking, validity shall describe whether a safety test evaluates what it should evaluate. Objectivity and reliability shall refer to errors caused by assessment assumptions that are subjective and uncertain, respectively. However, this characterization needs to be refined because it refers to errors directly and without further distinction, which is problematic for the following reasons. Errors are deviations from true values. For many aspects of a safety assessment, error quantification should be difficult either because the true state and behavior of the system is unknown or because statistical inferences about errors are not available. Referring to errors is also problematic because not all errors reduce the epistemic probability of operationalizing premises. For example, systematic overestimations of safety are tolerable when demonstrating insufficient safety. For this reason, test quality will not be defined with regard to errors but with regard to the probability of operationalizing premises. We will first consider, how this can be done for the reliability of a safety test.

Assessment results usually vary in the given range of uncertainty. This indicates how the given uncertainties might affect long-term protection. This piece of information becomes important when inferences about long-term protection are drawn from the result of safety tests (which is what operationalizing premises do). A compliance test, for example, may state that a site is safe, but it may be uncertain as to whether the test has neglected important repository processes. This uncertainty opens up the possibility that the level of long-term protection is insufficient, which would render the operationalizing premise untrue. Compliance tests can avoid this by implementing a strategy of conservatism. The ability to prevent operationalizing premises from being falsified by uncertainties is, therefore, a quality of the safety test, namely, its reliability.

Following this example, the validity, objectivity and reliability of safety tests are defined in the following way.

1. The *validity of a safety test* shall be the epistemic probability that the following propositions are true:
   a. The operationalizing premise of the rejection argument is not false solely because *the test does not measure what it should measure* according to conceptual requirements (for concept violation tests) or to the operational definition of the level of long-term protection (for all other tests).
   b. The part of the system understanding that is needed to understand how long-term protection is affected by deviations from what should be measured is sufficiently adequate.
2. The *objectivity of a safety test* shall be the epistemic probability that the following propositions are true:
   a. The operationalizing premise of the rejection argument is not false solely because of *subjective assumptions.*
   b. The part of the system understanding that is needed to understand the effect of subjective assumptions on long-term protection is sufficiently adequate.
3. The *reliability of a safety test* shall be the epistemic probability that the following propositions are true:
   a. The operationalizing premise of the rejection argument is not false solely because of *the given uncertainties.*

b. The part of the system understanding that is needed to understand the effect of uncertainties on long-term protection is sufficiently adequate.

2.9 Epistemic probability of operationalizing premises

We conclude the model description by defining the epistemic probability of operationalizing premises, which is needed to decide whether a rejection argument is sufficiently sound.

The *epistemic probability of an operationalizing premise* shall be the epistemic probability that the following propositions are true.

1. The operational definition of the level of long-term protection is sufficiently accurate.
2. The associated safety test is sufficiently accurate, that is, sufficiently valid, objective and reliable.

It is outside the scope of this model to explain how epistemic probabilities and sufficient probabilities can be ascertained (see discussion). For this reason, the model cannot offer a step-by-step guide to justifying site rejection decisions. Nevertheless, the it explains the nature and structure of the justification and leads to the following practical conclusions.

3 Practical conclusions

In this chapter, practical conclusions for site selection and safety assessment are presented. Please note that these conclusions draw on additional information on the practice of site selection and long-term safety assessment. Hence, they are not derived from the model alone.

3.1 Dealing with uncertainties

Rejecting sites whose characteristics are uncertain involves an increased risk of rejecting a safe or even the safest site. The acceptability of this risk is probably higher if promising sites remain.

There are different ways of accounting for uncertainties. For example, it is possible to ignore uncertainties and to base decisions on the apparent level of safety. In the presented model, this can be realized by lowering the expectation as to how reliable safety tests must be in order to be sufficiently reliable. The rejection arguments will then remain sound and the reduced reliability requirements will transparently display that the priority of safety targets has been reduced.

Alternatively, uncertainties could be interpreted as safety deficits, which is a risk-averse approach. A similar method was chosen by Fischer-Appelt and Baltes (2010) with regard to robustness deficits. However, the approach cannot be justified by the presented model. Interpreting subjective uncertainties as safety deficits could impair the objectivity of safety tests (according to the definition of 'objectivity' given in chapter 2.8, which is tolerant towards certain forms of subjectivity). Moreover, there is no causal relationship between epistemic uncertainties and the system's level

of long-term protection, which would render the latter's operational definition inaccurate.

3.2 Non-compliance tests

It was shown that non-compliance tests need to be anticonservative. In practice, non-compliance tests probably take the form of exclusion criteria. For this reason, it must be explained what an anticonservative exclusion criterion is.

An exclusion criterion shall be understood as a condition under which a site can be rejected. Often, it takes the form $i_n > c_n$ or $i_p < c_p$, where i_n and i_p are indicators for negative and positive safety factors, respectively, and c_n, c_p are constants that represent critical indicator levels. For example, i_p could be a barrier thickness and c_p a critical thickness; or i_n a could be a seismicity indicator and c_n a critical level of seismicity. Usually, the indicators i_n, i_p can be evaluated without comprehensive site characterization and scenario development.

Non-compliance test are not accurate enough if they exclude sites that might provide a sufficient level of long-term protection. An exclusion criterion must, therefore, overestimate the effect of positive safety factors that can falsify a statement of insufficient safety. Due to this (anticonservative) overestimation of safety, the critical indicator levels c_n or c_p must be raised or lowered, respectively, to ensure a high positive predictive value. This again reduces the number of sites that can be rejected by the test. Hence, the applicability of exclusion criteria decreases with increasing reliability. Reformulating exclusion criteria to let them reject as many sites as possible threatens their reliability and, thus, their capability of justifying site rejections.

3.3 Compliance tests

It might be necessary to assess certain sites so conservatively that their safety cannot be demonstrated. For sites of that kind, safety demonstrations require that conservatism is reduced by advancements in research, assessment methodology or site characterization. Whether there will be such advancements in the future is difficult to tell in the early phases of a long-lasting site selection procedure. Consequently, the truth of the rejection argument's second premise (chapter 2.3.2) is difficult to establish. Compliance test, therefore, should not be effective tools for site rejection in the early phases of a long-lasting site selection procedure.

Compliance tests are usually based thorough, site-specific and conservative assessments. Although it is not necessary for conservatism, it should be a reasonable practical requirement that the underlying assessments cover all important negative safety factors and detrimental scenarios that could falsify a statement of sufficient safety. An assessment of impossible evolutions is tolerable if it increases conservatism.

3.4 Ranking tests

It was mentioned that the accuracy of ranking tests does not depend on system properties and assessment errors that are shared by the compared sites. Increasing site dissimilarity should, therefore, increase the difficulty of safety rankings.

If sites are so dissimilar that they have to be assessed with differing scenarios (where the probability of a scenario shall be considered a scenario feature), the assessment inaccuracies caused by the assessment of impossible evolutions may differ too. Consequently, these inaccuracies cannot be ignored when the sites are compared. Unfortunately, they cannot be quantified because misperceptions of system possibilities are usually not detectable. For this reason, the accuracy of the ranking test remains unknown. This leads to the important conclusion, that cross-conceptual safety rankings cannot use indicators that require a comprehensive scenario development to predict possible evolutions of the repository system. Indicators of this type shall be called 'predictive indicators' and the dose indicator is an important example.

Cross-conceptual safety rankings, however, may be possible if they use indicators that can be evaluated without predicting possible evolutions (although the evaluation might refer to certain scenarios). Indicators of that kind shall be called 'non-predictive' and they are much less affected by misperceptions of system possibilities. Examples of non-predictive indicators are *diffusivity*, *host rock thickness* and *groundwater age*.

Whether the safety function indicators of a multi-barrier concept are non-predictive indicators, depends on their formulation. For example, *canister thickness* is a predictive indicator if it requires a prediction of how thickness will possibly be reduced by corrosion in the course of time. In contrast, it is not a non-predictive indicator if it refers to the safety margins of the canister layout, which can be determined without predicting future evolutions.

A drawback of non-predictive indicators is their remoteness from future exposures. For example, it should be difficult to determine how much the thickness of the host rock and the thickness of the canister should be weighted to reflect their physical importance for future exposures (without modelling possible evolutions). Consequently, the appropriateness of the weighting must be an operationalizing postulate.

In conclusion, the sources of inaccuracy appear to differ between predictive and non-predictive indicators: While the predictive indicators suffer from inaccurate safety tests, non-predictive indicators involve an inaccurate operationalization. Safety rankings fail if decision-makers do not accept these inaccuracies.

Safety rankings should not achieve different results if sites are compared pairwise and if the succession of pairwise comparisons is changed. Different results would indicate that the ranking test does not implement a transitive safety order. In this case, the ranking test is not accurate enough.

Cross-conceptional safety rankings will not be possible if decision-makers hold the view that different safety concepts cannot be compared in principle. This view is possible and not contradicted by the existence of a regulatory minimum level of safety. The reason is that the regulatory minimum level of safety may be concept-specific. In this case, repositories with different safety concepts are *sufficiently* safe at this level but not necessarily *equally* safe.

In conclusion, there can be indications that safety rankings are not accurate enough. The absence of such indications, however, does

not imply that safety rankings are sufficiently accurate. In this case, test accuracy becomes a matter of subjective judgement. The model expresses this by defining the accuracy of the safety test as an epistemic probability.

Inaccurate safety rankings may be perceived as a failure of safety assessment. Nonetheless, they might increase the efficiency of a site selection procedure by opening ways for other types of rejection arguments that do not require site characterizations. They might, instead, refer to other relevant aspects, such as conceptual requirements, operational safety or technical feasibility.

The model indicates that site similarity is important for accurate ranking tests. For this reason, cross-conceptual ranking tests might be facilitated by modifying existing safety concepts in order to increase site similarity and simplify safety rankings. Clay and crystalline concepts, for example, focus on different barriers. While the first utilizes the advantages of a low-permeability clay barrier with high sorption capacity, the latter concentrates on technical barriers with high performance. It might be possible to construct a combined safety concept by combining the geological barrier of the clay concept with the technical barriers of the crystalline concept. The similarity between the original concepts and the combined concept should then reveal a safety order that indicates a superior safety of the combined concept. The original concepts could then be rejected at the price of a more complex and expensive combined concept and this might increase the efficiency of the site selection procedure. Of course, the feasibility and long-term safety of such a concept must be demonstrated first. Moreover, the decision target of resource limitation can be a reason for not pursuing a combined concept. It is, therefore, not possible to derive a recommendation for this approach from the justification model.

3.5 Concept violation tests

Safety-related conceptual requirements considerably increase the scope of safety-related reasons for site rejections. In principle, every exclusion criterion that should turn out to be insufficiently reliable as a non-compliance or suboptimality test can be saved by repurposing it to a concept violation test.

Conceptual requirements can be used to reject certain safety concepts as a whole. For example, safety concepts could be rejected because they involve certain risks or types of uncertainties with which decision-makers feel uncomfortable. Such preferences can be elicited in an early stage of a site selection procedure to increase the efficiency of the site selection procedure.

The use of conceptual requirements is restricted by the fact that the requirements need to be consensual. Decision-makers might also want to limit the number of conceptual requirements because they do not prevent the rejection of sites that provide a high level of long-term protection.

4 Discussion

The difficulty of establishing an arguable connection between indicator values and long-term protection is a fundamental problem of site selection procedures if long-term protection is related to future exposures. This problem might not be obvious at first sight.

For example, the permeability of the host rock appears to be a reasonable indicator because increasing permeability should also increase future exposures. The indicator can therefore be used for comparing two sites that only differ with regard to this parameter. For this application, it is sufficient to know that permeability and future exposures are positively correlated.

Yet there are situations in which we need to know more about the relation between indicator values and future exposures. Regulators might, for instance, want to formulate an exclusion criterion based on the permeability of the host rock. But what permeability value would imply inacceptable future exposures? This is difficult to tell. To circumvent this difficulty, the criterion must prescribe a high permeability value, which again reduces the number of sites that will be excluded. Or take the example that decision-makers might want to compare sites that differ with regard to more than one indicator: Can they really find out, to what degree the individual indicators contribute to long-term protection physically? In these examples, it matters that the operationalization and assessment of future exposures is inaccurate. It is now necessary to understand how the justification of site rejections is affected by these inaccuracies.

The presented justification model clarifies this point. It shows that rejection arguments are tolerant towards specific types of errors. The model also points out that the justification gap between indicators and future exposures can be closed by operationalizing postulates, which do not claim to be true, but only to be acceptable. These postulates are indicator-specific and their acceptability must be ascertained.

The model is consistent with the notion that operationalizations and assessments of long-term protection aim at *predicting* or *measuring* the level of long-term protection or its boundaries. The general term 'accuracy' has been introduced to describe the suitability for making such predictions or measurements. The model specifies the meaning of this general term with regard to the accuracy of the operationalization and the accuracy of safety tests, both of which depend on the rejection argument.

Although the justification model was constructed under the assumption that long-term protection is related to future exposures, the model can also handle the assumption that long-term protection is defined in relation to the fulfillment of safety functions, i.e., to the quality of the measures taken to prevent future exposures. In this case, operationalizing postulates are not needed and showing that the operationalization of long-term protection is sufficiently accurate is a trivial task.

The OECD-NEA (2012) stated that 'it is also commonly understood that safety assessments are analyses that cannot and do not constitute absolute proof of safety, but efforts are made to design and conduct these analyses such that a high confidence in their results is achieved'. The presented model goes a step further by specifying how these analyses must be designed and conducted in order achieve confidence and on what this confidence depends. By doing so, the model displays that absolute proof is not required if the justification draws on subjective beliefs, for example, in the adequateness of the system understanding or in the validity, objectivity and reliability of safety tests. Nevertheless, the model does not clarify all details of justification, in particular with regard to inconsistencies of the system understanding. Future research should investigate which inconsistencies are tolerable and how indicators

and aggregation rules can be justified by an inconsistent system understanding.

The justification model operates with *sufficient* degrees of belief. It would have been possible, instead, to connect to evidence theory (Zhang and Jiang, 2021) and consider how degrees of beliefs combine. Yet this was not undertaken for the sake of clarity and because there seem to be more urgent practical problems at hand: The difficulties of measuring degrees of beliefs and of agreeing on sufficient degrees of belief.

The underlying problem is that degrees of belief are probably not uniform, i.e., not fully determined by the given body of knowledge. If beliefs are not uniform, both credulity and false beliefs are a matter of concern. False beliefs may form independently from or in contradiction to scientific evidence, for example, if stakeholders find scientific arguments too fragmentary or too complex to follow. False beliefs may even be caused by political ideology and affiliation, which, for instance, can influence the perception of uncertainties (Broomell and Kane, 2017). This shows that the justification of site rejection decisions cannot be understood by natural science alone. It requires interdisciplinary research. It also makes clear that rational site selection procedures should include processes of knowledge transfer and scientific review to reduce the probability of false beliefs. Also, credulity should be avoided by adopting a risk-averse, critical and self-critical attitude.

Aiming at uniform degrees of belief might also be a guiding principle when presenting modelling results and their uncertainties. It would then not be important how much uncertainties are disclosed, but whether the experts' trust in the modelling results is truthfully conveyed to the audience.

Let us turn to the role of safety assessments in site selection procedures. According to the justification model, safety assessments perform a double role: They must improve the system understanding as well as support the operationalizing premises of rejection arguments. For models that aim at improving system understanding, similarity with reality should be an important aspect of their representational nature. In contrast, models that are used for constructing compliance tests need not be similar to the real world if they only aim at being sufficiently conservative to support the operationalizing premise. It suffices here if 'the possibility to connect (i.e., to draw a cogent semantic relationship) between the model and the real-world is warranted by how the model is constructed' (Boon, 2020). The mentioned semantic relationship is given by the material implication of the operationalizing premise and the model's purpose is to create a sufficient degree of belief in the truth of that implication. On this account, the requirements for safety assessments depend on the assessment purpose.

The assessment purpose also determines the value of conservatism. Conservative assessments can obscure the true behavior of the repository system (Röhlig, 2024), which is problematic if the assessment's purpose is to develop system understanding. However, conservatism is not problematic for compliance tests that only aim at supporting the operationalizing premise of a rejection argument (assuming that the system understanding is already sufficiently adequate). For this reason, there is no general answer to the question of how conservative safety assessments should be.

The assessment purpose also controls the relevance of uncertainties as sources of inaccuracy. Uncertainties are often considered to be relevant to repository safety (Hoyer et al., 2021; Eckhardt, 2024), but this is only partially correct because the safety of a system does not depend on epistemic uncertainties, i.e., on missing knowledge. More precisely, uncertainties are relevant to the *assessment* of safety. How relevant they are, depends on the purpose of the assessment. There are uncertainties that are relevant to a certain operationalizing premise, but irrelevant to another. Moreover, uncertainties may be of minor relevance if the safety assessment is not designed to support an operationalizing premise but to generate ideas about how a repository system might behave. For this reason, the relevance of uncertainties cannot be evaluated without accounting for assessment purposes.

We now turn to the differences between site selection procedures and licensing procedures. One difference is that safety demonstrations for licensing procedures require conservative safety assessments, whereas justifications for site rejections mostly call for anticonservatism. Another distinguishing feature is that licensing procedures require reliable safety statements, whereas site selection procedures need to operate with less reliable safety statements if the resources for site characterization are limited. Consequently, site selection procedures must make trade-offs between safety targets and other targets, such as a reasonable limitation of resources. The presented justification model can account for such trade-offs.

An important practical conclusion from the model is that consensual conceptual requirements regarding the conditions under which long-term protection is achieved can be powerful tools for justifying site rejections. For example, a site might be rejected if it does not meet the requirement of radionuclide containment (even if it provided a sufficient level of protection). Although requirements of that kind can conflict with safety maximization targets, they still relate to a safety dimension and, therefore, provide strong rejection arguments. They might also be used to reject certain safety concepts as a whole.

The model also shows that comparisons of long-term protection (i.e., ranking tests) are sensitive to the inaccuracies of operationalization and safety assessment. Increasing site dissimilarity aggravates the problem. For this reason, it may not be possible to compare sites that differ with regard to host rock and safety concept. The model points out under which conditions this is to be expected. Firstly, there might not be any suitable indicators for such a comparison. Predictive indicators that require a comprehensive scenario development, like *dose* or *degree of containment,* are unsuitable in principle due to their unknown accuracy, which is caused by likely misperceptions of system possibilities. Non-predictive *i*ndicators, which do not require a comprehensive scenario development, do not suffer from this problem. However, they too can be unsuitable if decision-makers do not accept their inaccuracy of operationalization. Secondly, cross-conceptual safety comparisons are not feasible in principle if decision-makers claim that different safety concepts cannot be compared. In this case, no indicator will be able to establish a safety order.

Hence, there can be situations in which it is not possible to rank sites according to their level of long-term protection. Although this may be disappointing from the perspective of safety assessment, it can increase the efficiency of site selection procedures by opening the way for alternative comparison criteria that do not require extensive site characterizations, such as *fulfillment of conceptual requirements*, *technical feasibility* or *required resources*. In other words, site selection procedures may benefit from foundering safety assessments.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

MN: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review and editing.

Funding

Acknowledgments

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Anelli, A., Harabaglia, P., and Vona, M. (2025). Determining the location of the national repository of Italian radioactive waste: a multi-risk analysis approach. *infrastructures* 10, 22. doi:10.3390/infrastructures10010022

Bilgilioğlu, S. S. (2022). Site selection for radioactive waste disposal facility by GIS based multi criteria decision making. *Ann. Nucl. Energy* 165, 108795. doi:10.1016/j.anucene.2021.108795

Boon, M. (2020). The role of disciplinary perspectives in an epistemology of scientific models. *Eur. J. Philosophy Sci.* 10, 31. doi:10.1007/s13194-020-00295-9

Broomell, S. B., and Kane, P. B. (2017). Public perception and communication of scientific uncertainty. *J. Exp. Psychol. General* 146, 286–304. doi:10.1037/xge0000260

BT [Deutscher Bundestag] (2023). *Gesetz zur Suche und Auswahl eines Standortes für ein Endlager für hochradioaktive Abfälle: StandAG*. BGBl. I (Bundesgesetzblatt Teil I).

Chuaqui, R. (1991). *Truth, possibility and probability: new logical foundations of probability and statistical inference*. Amsterdam: Elsevier Science Publishers B.V. (North-Holland).

Department for Communities and Local Government (2009). *Multi-criteria analysis: a manual*. London.

Eckhardt, A. (2024). "Wie viel Ungewissheit ist akzeptabel? Beurteilung von Ungewissheiten in verschiedenen Entscheidungssituationen auf dem Entsorgungsweg," in *Entscheidungen in die weite Zukunft: Ungewissheiten bei der Entsorgung hochradioaktiver Abfälle*. Editors A. Eckhardt, F. Becker, V. Mintzlaff, D. Scheer, and R. Seidl (Wiesbaden: Springer Fachmedien Wiesbaden), 207–228.

Ewing, R. C., and Grambow, B. (2025). Final thoughts: the fragile connection of safety and science in the geological disposal of radioactive waste. *Bull. Atomic Sci.* 81, 48–52. doi:10.1080/00963402.2024.2439761

Finsterle, S., and Lanyon, B. (2022). Pragmatic validation of numerical models used for the assessment of radioactive waste repositories: a perspective. *Energies* 15, 3585. doi:10.3390/en15103585

Fischer-Appelt, K., and Baltes, B. (2010). *Abwägungsmethodik für den Vergleich von Endlagersystemen in unterschiedlichen Wirtsgesteinsformationen - Anleitung zur Anwendung der Abwägungsmethodik: Abschlussbericht zum Vorhaben 3607R02589 VerSi "Evaluierung der Vorgehensweise*, GRS-A-3536. Cologne: Gesellschaft für Anlagen- und Reaktorsicherheit gGmbH (GRS).

Gutberlet, D. (2015). *Can multi-criteria analysis models support the site selection for a repository for heat-generating waste?* Mining Report, 151, 188–195.

Heiermann, M., and Olszok, V. (2024). Transdisciplinary research on the safety case for nuclear waste repositories with a special focus on uncertainties and indicators. *Front. Nucl. Eng.* 3, 1414964–2024. doi:10.3389/fnuen.2024.1414964

Hoyer, E.-M., Luijendijk, E., Müller, P., Kreye, P., Panitz, F., Gawletta, D., et al. (2021). Preliminary safety analyses in the high-level radioactive waste site selection procedure in Germany. *Adv. Geosci.* 56, 67–75. doi:10.5194/adgeo-56-67-2021

IAEA [International Atomic Energy Agency] (2011). *Disposal of radioactive waste*. Vienna: IAEA.

IAEA [International Atomic Energy Agency] (2012). *The safety case and safety assessment for the disposal of radioactive waste*. Vienna: IAEA.

IAEA [International Atomic Energy Agency] (2022). IAEA nuclear Safety and security glossary, 2022 (interim) edition: terminology Used in nuclear safety, nuclear security, radiation Protection and Emergency Preparedness and Response. *2022 (interim) edition*. Vienna: IAEA.

ICRP [International Commission on Radiological Protection] (2007). *The 2007 recommendations of the international commission of radiological protection*. Elsevier.

Kelly, D., and Hutchins, D. (2021). *The art of reasoning: an introduction to logic*. W. W. Norton and Company.

Kuhlman, K. L., Bartol, J., Carter, A., Lommerzheim, A., and Wolf, J. (2024). Scenario development for safety assessment in deep geologic disposal of high-level radioactive waste and spent nuclear fuel: a review. *Risk Anal.* 44, 1850–1864. doi:10.1111/risa.14276

Liebscher, A., Borkel, C., Jendras, M., Maurer-Rurack, U., and Rücker, C. (2020). Towards best possible safety – current regulatory research for the German site selection process for high-level radioactive waste disposal. *Adv. Geosci.* 54, 157–163. doi:10.5194/adgeo-54-157-2020

Madeira, J. G., Alvim, A. C. M., Martins, V. B., and Monteiro, N. A. (2016). Selection of a tool to decision making for site selection for high level waste. *EPJ Nucl. Sci. Technol.* 2, 6. doi:10.1051/epjn/e2015-50039-x

Mallants, D., Bourdet, J., Camilleri, M., Crane, P., Delle Piane, C., Deslandes, A., et al. (2024). An assessment of deep borehole disposal post-closure safety. *Nucl. Technol.* 210, 1511–1534. doi:10.1080/00295450.2023.2266609

Merkhofer, M. W., and Keeney, R. L. (1987). A multiattribute utility analysis of alternative sites for the disposal of nuclear waste: decision analysis. *Risk Anal.* 7, 173–194. doi:10.1111/j.1539-6924.1987.tb00981.x

Müller, H. R., Blechschmidt, I, Vomvoris, S., Vietor, T., Alig, M., and Braun, M. (2024). Status of the site investigation and site selection process for a deep geological repository in Switzerland. *Nucl. Technol.* 210, 1740–1747. doi:10.1080/00295450.2023.2262298

NAGRA [Nationale Genossenschaft für die Lagerung radioaktiver Abfälle] (2024). *Qualitative Bewertung für den sicherheitstechnischen Vergleich in Etappe 3 des Sachplans geologische Tiefenlager: NAB 24-23 Rev*, 1. Switzerland: Wettingen.

OECD-NEA [Organisation for Economic Co-operation and Development - Nuclear Energy Agency] (2002). *The handling of timescales in assessing post-closure safety of deep geological repositories: workshop proceedings, Paris, France 16-18 april 2002*. Paris, France: OECD-NEA.

OECD-NEA [Organisation for Economic Co-operation and Development - Nuclear Energy Agency] (2012). *Methods for safety Assessment of geological disposal Facilities for radioactive waste: Outcomes of the NEA MeSA initiative*. Issy-les-Moulineaux, France: OECD-NEA.

OECD-NEA [Organisation for Economic Co-operation and Development - Nuclear Energy Agency (2013). *The nature and purpose of the post-closure safety cases for geological repositories*. Paris, France: OECD-NEA.

Okasha, S. (2003). Underdetermination, holism and the theory/data distinction. *Philosophical Q.* 52, 303–319. doi:10.1111/1467-9213.00270

Oreskes, N., Shrader-Frechette, K., and Berlitz, K. (1994). Verification, validation, and confirmation of numerical models in the Earth sciences. *Science* 263, 641–646. doi:10.1126/science.263.5147.641

Papafotiou, A., Li, C., Zbinden, D., Hayek, M., Hannon, M. J., and Marschall, P. (2022). Site selection for a deep geological repository in Switzerland: the role of performance assessment modeling. *Energies* 15, 6121. doi:10.3390/en15176121

Petraš, J. C. (1997). Ranking the sites for low- and intermediate-level radioactive waste disposal facilities in Croatia. *Int. Trans. Operational Res.* 4, 237–249. doi:10.1016/S0969-6016(97)00003-8

Röhlig, K.-J. (2024). "Ungewissheiten bezüglich der Langzeitsicherheit von Endlagern: Qualitative und quantitative Bewertung," in *Entscheidungen in die weite Zukunft: Ungewissheiten bei der Entsorgung hochradioaktiver Abfälle*. Editors A. Eckhardt, F. Becker, V. Mintzlaff, D. Scheer, and R. Seidl (Wiesbaden: Springer Fachmedien Wiesbaden), 253–281.

Schwenk-Ferrero, A., and Andrianov, A. (2017). Nuclear waste management decision-making support with MCDA. *Sci. Technol. Nucl. Installations* 2017, 1–20. doi:10.1155/2017/9029406

Taji, K., Levy, J. K., Hartmann, J., Bell, M. L., Anderson, R., Hobbs, B., et al. (2005). Identifying potential repositories for radioactive waste: multiple criteria decision analysis and critical infrastructure systems. *Int. J. Crit. Infrastructures* 1, 404–422. doi:10.1504/IJCIS.2005.006684

Tosoni, E., Salo, A., and Zio, E. (2018). Scenario analysis for the safety assessment of nuclear waste repositories: a critical review. *Risk Anal.* 38, 755–776. doi:10.1111/risa.12889

Turner, J. P., Berry, T. W., Bowman, M. J., and Chapman, N. A. (2023). Role of the geosphere in deep nuclear waste disposal – an England and Wales perspective. *Earth-Science Rev.* 242, 104445. doi:10.1016/j.earscirev.2023.104445

Valentin, J. (1997). *Radiological protection policy for the disposal of radioactive waste: ICRP Publication 77*. Oxford, UK: Pergamon Press.

Valentin, J. (1998). *Icrp publication 81: radiation protection recommendations as applied to the disposal of long-lived solid radioactive waste*. Oxford, UK: Pergamon Press.

Vigfusson, J., Maudoux, J., Raimbault, P., Röhlig, K.-J., and Smith, R. E. (2007). *European Pilot Study on The Regulatory Review of the Safety Case for Geological Disposal of Radioactive Waste Case Study: Uncertainties and their Management. Gesellschaft für Anlagen-und Reaktorsicherheit mbH (GRS); Hauptabteilung für die Sicherheit der Kernanlagen (HSK); Federaal Agentschap voor Nucleaire Controle – L'Agence fédérale de Contrôle nucléaire (FANC – AFCN); Autorité de Sûreté Nucléaire (ASN)*; Environment Agency (EA). Brussels: Föderale Agentur für Nuklearkontrolle (FANK).

Zhang, Z., and Jiang, C. (2021). Evidence-theory-based structural reliability analysis with epistemic uncertainty: a review. *Struct. Multidiscip. Optim.* 63, 2935–2953. doi:10.1007/s00158-021-02863-w