Table C3: The length of historical record required to estimate return period events with 95% and 80% reliability.

Return Period (Years)	Length of record in years required to deliver reliability of return period estimate	
	95% reliable	80% reliable
2.33	40	25
10	90	38
25	105	75
50	110	90
100	115	100

From Lee and Jones (2004) after Benson (1960).

With sufficient data it may be possible to formulate Frequency vs Magnitude curves to summarise the data and gain a better understanding of the overall process and associated frequencies. (For example, refer Moon, Wilson & Flentje 2004, and MacGregor *et al.,* 2007).

iii) Evaluation of Rainfall

Statistical evaluation of rainfall data is relatively easy to perform using computer spreadsheets. These statistics can be related to the incidence of landslides. An example is given in MacGregor *et al.* (2007).

Consideration has to be given to possible trigger thresholds which may relate to rainfall, either in the short term (minutes to hours) or the long term, such as antecedent rainfall over weeks to months. Usually, antecedent rainfall will be relevant where rising groundwater levels are seen as the main trigger, and this is frequently applicable for the larger landslides

In addition, there may be a conditional probability of the landslide event occurring during a given rainfall event, or the conditional probability related to the proportion of similar slopes that might be affected by a rainfall event. Such conditional probabilities may be evaluated by considering the proportion of slopes that have failed in a given rainfall event (based on the landslide inventory in conjunction with the rainfall analysis).

Use of simulation models which predict piezometric responses to rainfall events may assist with calibration and extrapolation to extreme rainfall events. However, these require long periods of records of rainfall and piezometric data, and even when this is available simulation is difficult. Fell *et al.* (1991) gives an example. Table C4 indicates the probability of different return period events occurring over different periods of time. It can be seen that the probability of having a low return period event (for example a 1 in 100 year event) over a relatively short monitoring period such as 5 years is quite low (4%). Thus such models and extrapolation will have obvious limitations but may still be a useful tool for understanding a particular scenario.

Table C4: Percentage probability of the N-Year event occurring in a particular period.

Number of years in period	N = Average return period in years							
	5	10	20	50	100	200	500	1000
1	20	10	5	2	1	0.5	0.2	0.1
5	67	41	23	10	4	2	1	0.5
10	89	65	40	18	10	5	2	1
30	99	95	79	45	26	14	6	3
60	>99.9	98	95	70	31	26	11	6
100	>99.9	99.9	99.4	87	65	39	18	9
300	>99.9	>99.9	>99.9	99.8	95	78	45	26
600	>99.9	>99.9	>99.9	>99.9	99.8	95	70	45
1000	>99.9	>99.9	>99.9	>99.9	>99.9	99.3	87	64

After Lee and Jones (2004).

The effects of 'climate change' may show that use of historical rainfall records has an implied limitation. However, at this stage the effect of climate change cannot be predicted. Some predict longer dry periods, whilst others are predicting higher intensity rainfalls. Since it may be that a changed rainfall pattern may in many cases increase the probability of landsliding, whilst dryer periods may decrease the probability for others, it is considered appropriate at this time not to attempt to adjust the assessed frequency for such changes.

iv) "Degree of Belief" or Subjective Probability

For many cases, the practitioner will have to rely on the "degree–of-belief" method or subjective probability. This will be necessary due to the lack of relevant information such as historical records and/or quantitative analysis of trigger events which would enable an objective assessment of event probabilities. The practitioner will have to make best estimates of frequency/likelihood from limited site data, using experience and broad knowledge of an area or other areas of similar slope form and geology.

Moon and Wilson (2004) provide a useful over-view to developing judgments on landslide likelihood. "*The necessary evidence on which judgments of landslide likelihood are based has to be assembled, understood and interpreted. This process involves developing a slope model that reflects a sound knowledge of how the slope was formed, how it behaved in the past and how it might behave in the future. The ability to build up such a model comes from knowledge of the slope and its surrounds, knowledge of similar slopes in similar environments, and a range of skills and knowledge bases that result from training and experience.*" Many useful references are cited.

Vick (2002) discusses the role of evidence and logical inference to subjective probability and engineering judgment. Although the assessed likelihood will be a subjective judgment, it should, like a bookmakers odds, be based on evidence (Moon and Wilson, 2004).

There are undoubted problems associated with use of "degree–of-belief" methods. The following presents a summary of the discussion in Lee and Jones (2004).

The main potential problems identified by Roberds (1990) are, in summary:

- *Poor quantification of uncertainty*, which may result in significant over estimates of likelihood where the slope forming process is ignored or misunderstood.
- *Poor problem definition*, as a result of the practitioner's experience and background, resulting in emphasis on one area or element of the slope at the expense of another.
- *Motivational bias* which may result in over optimistic or overly conservative assessments depending on the purpose of the assessment.
- *Cognitive bias* where the practitioner's judgment does not match the available facts.

The effects of these potential problems can be reduced or eliminated by techniques such as those of Lee and Jones (2004):

- "*Self assessment* where the rationale behind every judgment has to be well documented as required by the Practice Note. The same operator bias is likely to apply, but the documentation process clarifies the logic and results in a more defensible judgment.
- *Independent review or second opinion* which also should be well documented. This may still suffer from bias.
- *Calibrated assessment* where the practitioner's biases are identified and calibrated, and the assessment adjusted accordingly. The biases may be identified by peer group review or objectively by a set of experiments or questionnaires.
- *Probability encoding*, which involves the training of practitioners to produce reliable assessments of the probability of various events in a formal manner. This involves six stages:
    1. Training the practitioner to properly quantify uncertainty.
    2. Identifying and minimizing the practitioner's bias tendencies.
    3. Defining and documenting the item to be assessed in an unambiguous manner.
    4. Eliciting and documenting the practitioner's rationale for the assessment.
    5. Eliciting, directly or indirectly, the practitioner's quantitative assessment of uncertainty and checking for self-consistency. The practitioner's uncertainty can be established by determining the probability of various states through comparison with reference situations, such as poker hands, or by choosing between two lotteries (e.g. probability wheels or intervals) until indifference is achieved.
    6. Verifying the assessment with the practitioner and repeating the process if necessary."

Group consensus about a judgment is desirable but is achieved at increased cost and may not be economic. There may be significant differences of opinion between different practitioners. Where such differences of opinion are identified then they should be attempted to be resolved preferably in an open forum. The outcomes from this resolution process can be:

- *Convergence* to a common belief or assessment agreed to by all practitioners in the group.
- *Consensus,* where a single assessment can be determined but the assessment may not be the exact view of each individual. The consensus assessment may be a compromise derived from the individual assessments of group members but without the express agreement of the individuals concerned (forced), or the group may expressly agree to it for a particular purpose (agreed).
- *Disagreement.* Where convergence or consensus to a single assessment is not possible from the multiple assessments due to the major differences of opinion.

More detailed discussion of the above is presented in Lee and Jones (2004).

The Working Group considers that the Practice Note outlines "best practice" for self assessment where a "degree–of-belief" method is frequently adopted. However, it is anticipated that documentation of the assessment will include reference to known history and trigger events to help calibrate the judgment and provide defensibility.

The assessment of frequency should adopt the best means available given the nature of the landslides, circumstances of the geotechnical model, nature of triggering events and requirements of the risk assessment. Where few data are available, then estimates erring on the conservative side should be adopted to cover inherent uncertainty. More detailed studies may then be required to provide more reliable risk estimates.

In considering the circumstances of the particular assessment, the practitioner has to use best estimates from the available data when assigning likelihood (and consequences) values, but will inevitably be based on a subjective assessment of the practitioner's "belief" of the assessment. The assessment needs to consider range/uncertainty/sensitivity of the assessed values to establish confidence. The practitioner has to apply judgment, but must provide an explicit trail, or explanation, of logic applied to derive the best estimates adopted.

Stewart *et al.* (2002) discuss the RTA Guide to Slope Risk Analysis which provides a systematic procedure for LRM for roads based on defined ratings to derive an Assessed Risk Level. The companion paper (Baynes *et al.,* 2002) discusses the issues of accuracy and precision in use of the procedure by many practitioners on a large number of slopes. The methodology of the procedure is based on principles outlined above. Training in use of the system is required to help calibrate each practitioner and reduce bias. Audit procedures are used to derive consensus where necessary.

The state-of-the-art paper by Picarelli *et al.,* (2005) also provides a further overview and examples.

v) Event trees

Event trees enable the logical sequence of events to be considered in a structured manner. A suitable structured approach might, for example, consider for each scenario sequences such as likely trigger event, slope response, and consequence. An event tree can be used for complex scenarios.

The method has the advantage of enabling the logic adopted to be clearly shown together with each estimate of conditional probability, thereby providing clear documentation for review and appraisal.

This matter is discussed further in Lee and Jones (2004), and provides some examples where the method has been used. Hsi and Fell (2005) give an example where triggering by rainfall, over-taxing of a culvert and earthquake is modelled. Mostyn and Sullivan (2002) provides examples in relation to failure of fill embankments along a road. Hill *et al.* (2002) provides further discussion of issues associated with the principles of event trees.

5.4.2 Estimation of Annual Probability (Frequency) ($P_{(H)}$) of Each Landslide

a) Use best estimates for frequency estimates but consider range/ uncertainty/ sensitivity.

AGS (2000) acknowledged that assessment of frequency, or likelihood, is the most difficult part of the risk assessment process.

Assessment is particularly difficult at the medium to low frequency end (say 10^{-4} pa to 10^{-6} pa) because historic data based methods are not applicable. However, such values may still be appropriate by a combination of understanding the slope forming processes and logical elimination of other values. For some cases, such low frequency values may obviously be appropriate to hazards which could only occur over periods of geological time.

Experience has shown there is an inherent danger with Appendix G of AGS (2000), in that some practitioners assessed the likelihood solely based on the Descriptor. The Indicative Likelihood would then be adopted without due consideration. This procedure is incorrect as described below. An estimate of the probability should be made based on the best estimate of performance, trigger probabilities etc. and then the descriptor may be assigned accordingly.

Words such as "likely" can mean many different things to different people and in various contexts. The likelihood descriptors vary enormously in probability value between different publications as shown in the attached Table C5.

The qualitative terminology for Likelihood adopted for the Practice Note Appendix C is essentially the same as Appendix G, AGS (2000). The lowest category of likelihood has been revised to Barely Credible (from Not Credible).

The Descriptors are given to provide a consistent set of terms to assist the non-practitioner to interpret the assessed annual probability. In addition, the Descriptors provide a useful summary term for discussion purposes with due recognition of the inherent limitation of accuracy that is involved.