Executive summary

Subgroup 1: Management Strategy Evaluations (MSEs) and Retrospective Bias

Comparing MSEs with and without full assessments
Perceptions of the impact of harvest-control rules (HCRs) on the underlying population can change with different levels of approximation in the MSE, particularly when considering the longer term. These changes are not always obvious when considering only individual distributions of quantities of interest (e.g. SSB and landings yield). It is possible that, using an approximated MSE (e.g. omitting both the assessment and intermediate-year lag) could lead to perceptions of superior performance of one HCR relative to another in terms of summary statistics that would not be concluded if a full MSE were conducted.

Splitting survey series in response to retrospective patterns
In all cases examined, splitting the survey series produced fishing mortality rates in the population much closer to the target value than ignoring the source of the retrospective pattern in the original assessment. This conclusion held independent of the source of the retrospective pattern and whether it was a step change or gradual change. However, the best performance was found for models which most closely met the assumption underlying the split survey approach: a change in survey catchability or a sudden step change in process. Splitting of surveys is not recommended as a routine fix for all assessments exhibiting retrospective patterns: external information should be used to guide the decision about how to address an assessment with a retrospective pattern.

Analyzing and summarizing MSE outputs
Conducting an MSE can be approached in a two-step procedure. The first step would be done among the scientists, and would address the issue of model dimensionality—which factors are important, and how many levels and/or iterations should be retained for the ‘final’ MSE. The second step involves summarizing the information for managers and the general public, and this should be done graphically. All results of the MSE could be put into an appendix table for persons interested in the fine details.

Catch advice when different methods used to account for retrospective
Catch advice in the cases examined was relatively insensitive to the assumption made to address the retrospective patterns, being always lower than the original catch advice from the unadjusted base assessment as a consequence of the direction of the retrospective patterns.

A state-space fish stock assessment model applied to a dataset with a strong retrospective pattern
The use of random effects in state-space models for fisheries stock assessment appears to be a promising avenue for future research as a consequence of the limited number of parameters and quick run times to produce both point estimates and measures of uncertainty. These models can produce results without retrospective patterns in cases where traditional VPA assessments exhibited strong retrospective patterning.
Recommendations regarding retrospective patterns in stock assessment

WGMG considered the recommendations presented to the 2008 US Groundfish Assessment Review Meeting (GARM) Data Meeting regarding retrospective patterns in stock assessment. Several of these were accepted as appropriate, and are summarized in Appendix 3.

Subgroup 2: Uncertainty in stock assessment models

SURBA

The SURBA assessment method can provide unbiased estimates of population quantities when information is available on the relative catchability of the survey or tuning index. However, even in this unrealistically good situation, SURBA CIs can have poor simulated coverage properties. Most notably, CIs for total mortalities were too wide, and CIs for recruitments were too narrow. However, CIs for SSB were fairly reasonable.

SURBA was also sensitive to penalty function weights, especially when relative catchability is estimated. This particularly affected estimates of mortality at younger ages, and biomass and SSB estimates. We cannot provide any guidance on how to choose these weights in practice, other than trial and error. SURBA estimates of (relative) recruitment and total mortality at older ages seemed more reliable.

A random-effects approach is a more objective way to deal with controlling the variation in high-dimensional parameters. The variance of the random effects is analogous to the penalty weight, and these variances can be estimated (i.e. chosen objectively).

A state-space stock assessment model

New algorithms and software tools capable of optimizing a full state-space stock assessment model in minutes rather than in hours allow for the investigation of the frequentist properties of such models. It can be concluded that the simple confidence intervals for average fishing mortality and spawning-stock biomass are too narrow. It is recommended that more advanced ways of constructing confidence intervals (profile likelihood, or simulation based methods) are investigated.

More generally, state-space fish stock assessment models are worthy of further investigation, as they are able to separate process and observation noise and they avoid arbitrary smoothing parameters and ad-hoc weighting of different data sources.

Stock assessment models incorporating partial information about discards

The two stock assessment models presented are able to incorporate discards data (and other kinds of catch data, such as bycatch) that may be available in just some of the assessment years. The models produce stock assessments that incorporate all the available catch information, although at the same time, complete time-series of model estimates are obtained for discards (and any other components of the catch incorporated in the model, such as bycatch). The model estimates have associated confidence (in the maximum likelihood setting) or posterior probability (in the Bayesian setting) intervals and the stock assessment results also incorporate this uncertainty.

The finding that the model predictions are relatively robust against removal of part of the discards data relies entirely on the quality of the survey tuning indices. It appears that there is enough information in these to do a reasonable reconstruction of the discards. This is of course no guarantee that the same is true for all fish stocks. Also, the
reconstruction of the discards is only possible in years where survey tuning series exists. This may hamper reconstructing discards for very long time-series.

**Subgroup 3: Detecting changes in stock productivity**

The test applications implemented show that available change-detection methods can help to resolve a problem regularly encountered by assessment Expert Groups – that of deciding when the underlying recruitment state has change to a lower level. For both NS cod and NS haddock, the methods were fully consistent in the identification of breakpoints in recruitment series despite their distinct theoretical frameworks. Beyond the advantage of permitting an objective choice of breakpoints, the methods are also very simple to implement.