Benchmarking exploration predictions and performance using 20+ yr of drilling results: One company's experience

Kurt W. Rudolph and Frank J. Goulding

ABSTRACT

Historical exploration drilling results provide an opportunity to test the accuracy of geoscience interpretations and technologies by comparing predrill predictions to postdrill outcomes. This includes chance of success, success case recoverable hydrocarbon volumes, and individual reservoir parameters. Analyzing Exxon Mobil's conventional wildcat predictions versus results from 1994 to 2015 leads to the following insights. (1) Including all wells, risking and volume predictions were objective. Predrill predictions overall differentiated between high- and low-risk prospects and large and small volumes. However, individual wildcat volumes had significant uncertainty, with a lognormal distribution. (2) Prospect parameter predictions were also subject to considerable uncertainty. For example, net-to-gross error was ±20%. (3) Exploration play maturity strongly influenced performance. New play tests had a lower success rate but very large success case volumes. Chance of success increased and prospect success case volumes decreased with play maturity. For very mature plays, success rate decreased again. (4) Trap and seal failure accounted for about half of all dry holes. However, source, maturation, and migration are the most important risks for play tests and extensions. (5) Two seismic technologies were associated with success rate differences. Wildcats drilled based on three-dimensional seismic data had 10%–15% higher success rate than those based on two-dimensional data. Direct hydrocarbon indicator (DHI)–based prospects had about double the success rate of non-DHI prospects and were also overrisked. Although it can be misleading to use previous performance as an indicator of future results, benchmarking geoscience analysis with historical outcomes is useful to audit technical work, identify areas for improvement, and guide future predictions.

Manuscript received April 4, 2016; provisional acceptance May 27, 2016; revised manuscript received June 14, 2016; final acceptance June 28, 2016.
DOI:10.1306/06281616060

AUTHORS

Kurt W. Rudolph ~ *Retired, 202 Willard St., Houston, Texas 77006; kwrudolph@comcast.net*

Kurt W. Rudolph worked in a variety of assignments for Exxon Mobil for 34 years, including chief geoscientist from 2002 until his retirement in 2015. He is currently an adjunct faculty member of the University of Houston. He holds geology degrees from Rensselaer Polytechnic Institute (bachelor's) and University of Texas at Austin (master's). His interests include stratigraphy, seismic interpretation, and regional geology.

Frank J. Goulding ~ *ExxonMobil Exploration Co., 2777 Springwoods Village Parkway, Spring, Texas 77389; frank.j.goulding@exxonmobil.com*

Frank J. Goulding is chief geoscientist at ExxonMobil. He has worked in petroleum geoscience at ExxonMobil for 29 years in a variety of exploration, development, production, and research assignments. He graduated from the University of Calgary (geophysics) in 1987. His interests include seismic interpretation and response to hydrocarbon, stratigraphy, and production and development geoscience.

ACKNOWLEDGMENTS

The authors thank Exxon Mobil Corporation and the hundreds of company geoscientists and engineers who contributed indirectly and directly to this study over the course of 20+ years. Interim analyses by Emily Wall in 2001 and Martin Collier and Scott Johnsgard (ExxonMobil Exploration Co.) and Lonnie Smith (ExxonMobil Development Co.) in 2012 provided valuable additional insights. Early review of this paper by Kurt Steffen (ExxonMobil Development Co.), Lance Cook (XTO Energy, Inc.), and Kevin Biddle greatly improved the manuscript. IHS Markit permitted integration of their data into Figure 11. James Handschy, William Hill, and Michael Sweet provided expert editorial review before publication.

EDITOR'S NOTE

Color versions of Figures 1–7, 9–19, and 20 can be seen in the online version of this paper.

INTRODUCTION

Fundamental interpretation and analysis, based on geoscience principles, form the basis of the work model for identifying and characterizing exploration drilling prospects. Historical drilling results, based on both industry and proprietary experience, provide an additional means to test the ideas and techniques used by geoscientists during risking and assessment. Moreover, this type of analysis is a check on the fidelity of our technical predictions, including highlighting potential gaps, issues, or biases.

This paper summarizes findings from an ongoing effort within Exxon Mobil Corporation (hereinafter Exxon Mobil) to quantify predictions and outcomes on a parameter, volume, and risk basis. Most of the discussion is focused on exploration for conventional oil and gas resources. However, information on unconventional resource exploration, development, production, and industry drilling results are also used for comparison. Although not every data entry is complete for all wells, we used this information to test predrill assumptions against results on a global, business stage, resource type, or analog basis and to help better identify and evaluate opportunities by highlighting where there may be a tendency to overpredict or underpredict on a risk, volume, or parameter basis. Although previous performance is not necessarily a reliable predictor of future performance, the trends described herein have been largely stable over a period of more than 20 yr and are probably grossly representative of future overall results.

The complementary aspects of fundamental geoscience analysis and global to local calibrations are illustrated in Figure 1. Petroleum geoscientists make predictions about plays, prospects, reservoirs, and well performance. This is accomplished by applying concepts, tools, and workflows to standard subsurface data sets. Results, typically associated with the drilling of an exploration well, inform us of the efficacy of our technologies and our understanding of the geology in general. As we accumulate this information from several wells, it can be aggregated on a local and global basis. When the sample is large enough, such information can be used to calibrate our concepts and tools. Moreover, such statistics can be used as a benchmark for future predictions. It is extremely important that such data be placed in a firm geologic and business context, to ensure that the analog calibrations are relevant ("apples to apples").

ASSESSMENT METHODOLOGY

The assessment and risking procedures used by Exxon Mobil are generally similar to those used by most large oil and gas companies (Rose, 1999). For clarity, a review of the applied work process and nomenclature is provided below.

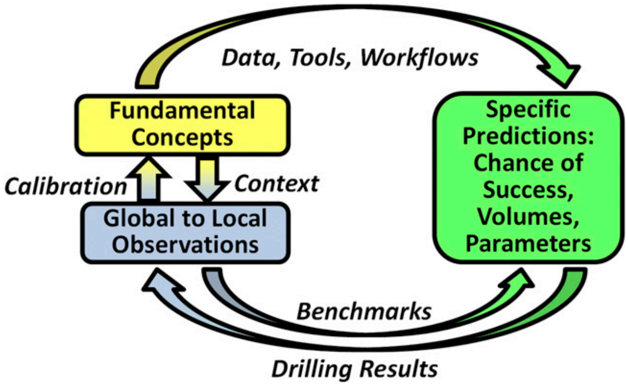

Figure 1. Generalized work model, illustrating the relationship of benchmarks, based on results (such as drilling outcomes), to petroleum geoscience analyses, which are generally grounded in fundamental concepts. These two complimentary approaches can improve predrill predictions.

Risking and volumetric assessment is done via consensus within the technical team integrating external advisors or experts. The team evaluates the best estimate and a range of scenarios for each parameter integrating offset control and analogs. This is further evaluated (and potentially modified) by a structured peer review by senior advisors. Many prospects cycle through several generations of risking and assessment as they mature. The final predrill estimates are captured in a prospect inventory database that is the source of predrill and postdrill information used in this analysis.

Risking is first done on a technical success basis (geologic chance of success [GCOS]) for individual reservoirs, with risk dependencies between multiple targets specified. Up to nine independent risk elements are estimated using a standardized risking matrix (White, 1993; Sykes et al., 2011). The chance of success for a reservoir target is the product of all these elements. Risking of each element is referenced to a specified geologic minimum volume that is set as a criterion for geologic success (White, 1993). The geologic minimum volume is typically much smaller than what would be indicative of economic success and is commonly in the range of 1–10 million oil-equivalent barrels, recoverable, for the hydrocarbon accumulation or field (Figure 2A). Prospects with multiple targets or segments are aggregated, with dependencies specified, to develop a composite chance of success. Unconventional prospects are risked against a well productivity minimum instead because a minimum field size is less meaningful.

The volumetric assessment is developed concurrently with risking, and a range of parameters and volume factors are described, including dependencies between factors and reservoir segments. These are simulated using Monte Carlo methods. One or more deterministic scenarios are commonly compared with the resulting aggregated probabilistic distribution and statistical measures such as the mean. The deterministic scenarios are then used for, among other things, engineering and economic analysis based on a notional development plan.

Economic chance of success (ECOS) is derived by applying a field volume threshold (economic minimum)

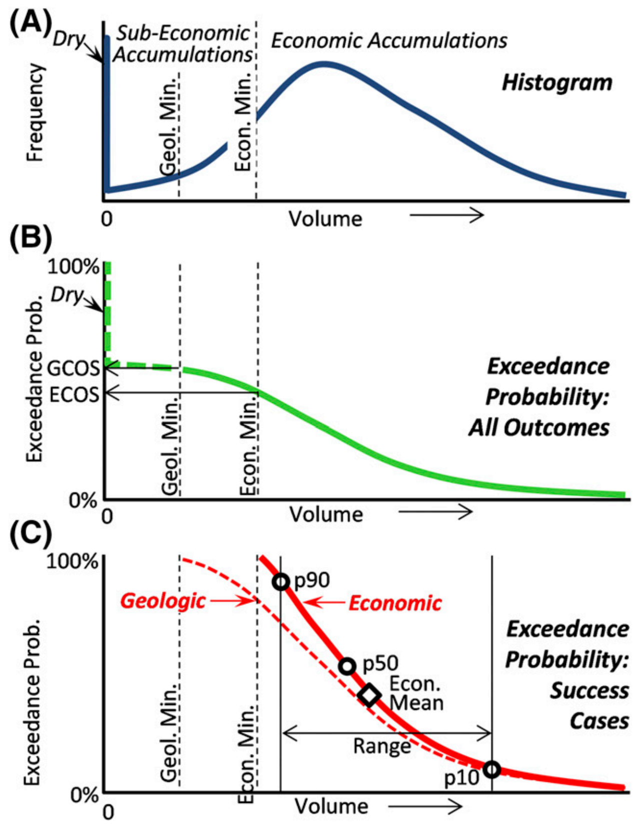

Figure 2. Schematic probability (p) distributions for an exploration prospect. (A) The frequency distribution produced via a Monte Carlo simulation of a predrill assessment model is more commonly converted to a (B) total exceedance probability curve that communicates risk and volumetric outcomes. (C) Successful trials can be redisplayed to illustrate the distribution of success cases. Outcomes above an economic threshold are selected to develop estimates of the economic chance of success (ECOS) (B) and the economic mean (C). Econ. Min. = economic minimum; GCOS = geologic chance of success; Geol. Min. = geological minimum; p10 = 10th percentile; p50 = 50th percentile or mean; p90 = 90th percentile; Prob. = probability.

to the probabilistic distribution of volumes (Figure 2A, B). Commodity (gas vs. oil), resource density, per-well productivity (for unconventional hydrocarbons), or other factors may also be applied to define the economic threshold. The predicted economic success cases eliminate interpreted noncommercial outcomes, and this truncated distribution has a lower chance of success and higher mean than the geologically risked curve (Figure 2B, C). The economic success case mean is recorded as the volumetric prediction for a prospect (Figure 2C). Moreover, the volume prediction for a portfolio of prospects (as on an annual drilling program basis) is based on the sum of the risked means: Σ_n (ECOS $\times$ economic mean).

As illustrated in Figure 2B, C, predrill volumetric assessments are commonly described by a cumulative frequency distribution, which describes the probability (y axis) of a prospect exceeding a given recoverable volume (x axis). Key statistical measures that help characterize the distribution are the mean, median (50th percentile [p50]), high-side (10th percentile [p10]), and low-side (90th percentile [p90]). We use the ratio of high side to low side (p10 to p90) as a simple measure of uncertainty, which is defined as the range of successful outcomes (Rose, 1987). Within this paper, this same type of display is also used to illustrate normalized volumetric predrill versus postdrill error (%) for a population of successful prospects.

DATABASE AND ANALYSIS

The primary data for the analysis are Exxon Mobil wildcat predictions and results from the period 1994–2015. A majority of the rank wildcats drilled by Exxon Corporation, Mobil Corporation, and Exxon Mobil are included and represent exploration drilling in 44 countries. Data from a minority (<20%) of wildcats, mostly in the 1994–1999 period, have not been recovered, and not all data fields are populated for the 500+ wells that are included in the study. Information on exploration appraisal wells and a large number of near-field wildcats (or outposts) are not included in this analysis. Unconventional wildcats are part of the study but are studied separately and given less treatment because of a smaller data set; less run-time; and more ambiguity in distinguishing exploration, appraisal, and exploitation.

Data parameters include predrill and postdrill estimates of geologic chance of success, economic chance of success, and recoverable volumes (postdrill chances are deemed as 0 or 1). For context, information on year drilled, exploration maturity, play stage, country, basin, play, seismic control, direct hydrocarbon indicator (DHI) support, geologic age and gross environment of deposition of the primary target, primary source type, trap style, and reason(s) for failure for dry holes are also included. Parametric predictions and results (reservoir net to gross, porosity, hydrocarbon saturation, and depth) were captured and evaluated for a smaller subset of wells. This information is only for the primary drilling objective if the prospect had multiple targets. Additional data sets on development projects, production drilling programs, field revisions, and industry wildcat results have been used for comparison purposes but are not described in detail herein.

Analysis was done by comparing predrill predictions to outcomes on a risk, volume, and parameter basis. Every wildcat is deemed a success or failure on a geologic (GCOS) and economic (ECOS) basis in the year of the well's completion. Economic successes have met internal economic criteria and have been entered into the company's resource base. Geologic successes exceed the geologic risking minimum. Analyses were performed on subsets of the data for elements such as predrill risk tier, play maturity, DHI support, depositional environment, etc. Subpopulations generally have more than 50 data points (and commonly many more) to avoid the pitfalls of small data sets as described by Peel and Brooks (2016). This is especially important when evaluating the performance-based risk for either very high or very low chance of success prospects.

A potential issue is that the data are not representative, because of significant dependencies between wells on a risk and volume basis. This is ameliorated by the fact that approximately 75% of wells are in proven plays, where prospect risk and volume considerations predominate over shared play risk and volume factors. For play tests and play extensions, higher risk and volume dependencies are expected; however, a high diversity of tests, with 68 basins tested by 145 wells, indicates that this is not a significant issue.

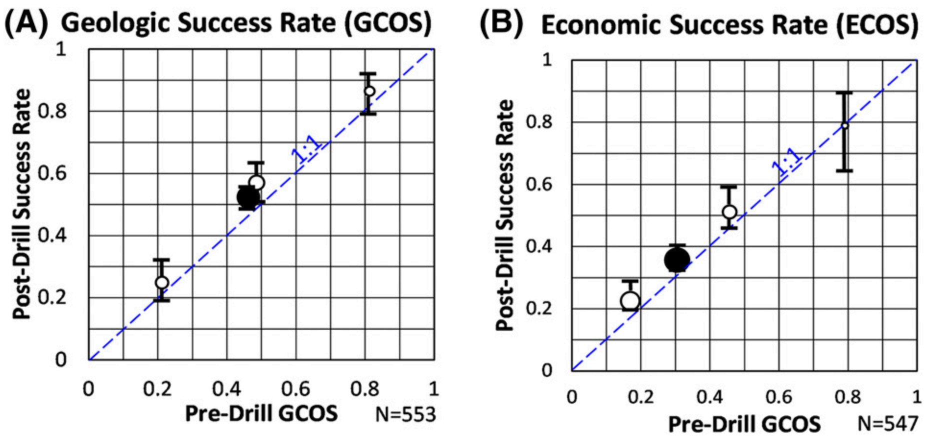

Figure 3. Comparison of predicted versus actual (A) geologic and (B) economic success rates for 1994–2015 conventional wildcats. Areas of the circles are proportional to sample size. The whiskers on actual success rates are the 95% confidence interval based on the binomial distribution. The total program delivered slightly higher success rates than predicted (black circles). The smaller open circles show precomparison versus post-comparison on a tiered predrill risk basis (0–0.33, 0.34–0.66, and 0.67–1.00), indicating a good discrimination of low- versus high-risk prospects. Nearly half the wells were in the middle risk tier for geologic chance of success (GCOS), but a majority were in the low chance of success tier for economic chance of success (ECOS). *N* = sample size.

RESULTS

Success Rates

Over the analysis period (1994–2015), risking for conventional wildcats has been generally accurate on both a geologic (GCOS) and economic (ECOS) basis (Figure 3). The predicted success rates were slightly pessimistic, with the actual success rate approximately 5% greater than estimated. Although a small difference, it is outside of the 95% confidence interval because of the large sample size. Moreover, subdividing the results by predrill risk tier illustrates very good discrimination of low from high risk (Figure 3). In terms of temporal trends, there are no large changes, although both the predicted and actual success rates have generally been higher since about 2002 (Figure 4). Relative to the 1994–2001 period, predicted success rates increased 6% and 3% for GCOS and ECOS, respectively. Corresponding actual success rates increased 10% and 13%, with underprediction of success being much greater in the 2002–2016 period (gap between predrill and actual lines in Figure 4). The reasons for the improved success rate in more recent drilling might be related to higher selectivity, improved technology, a changing portfolio, or other factors not recognized. This subject is revisited later in this paper.

Unconventional Wildcat Results

Risking of unconventional wildcats is a somewhat different topic: as described above, geologic and economic risks are estimated against criteria of per-well recovery thresholds. Approximately 120 unconventional wildcats were drilled by ExxonMobil Exploration Company over this period (Figure 5). A variety of resource types were targeted: tight gas, shale gas, shale oil, tight oil, and coal bed methane (heavy oil was excluded). Overall success rates are much higher for unconventional prospects (compare Figures 3 and 5), although the final economic disposition may be more uncertain. Hence, greater potential exists for the actual economic success rate to be modified in the future, as the well recovery and commercial outlook is revised.

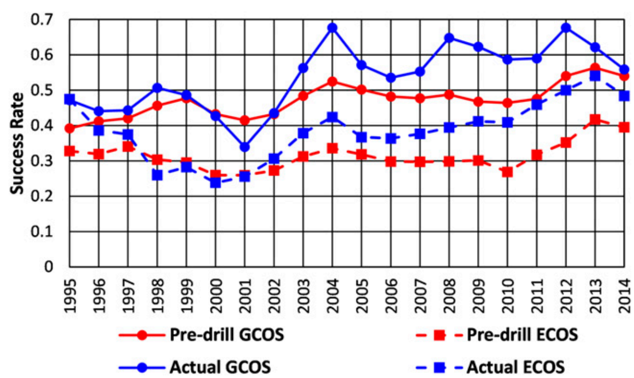

Figure 4. Success rate time trends for 1994–2015 Exxon Mobil conventional wildcats, with circles for geologic chance of success (GCOS) and squares for economic chance of success (ECOS) (predicted and actual). This chart uses a 3-yr moving average, with the horizontal axis plotted for the middle year. Although no dramatic change was noted, geologic and economic success has been somewhat higher for the last approximately 12 yr. Sample size = 553 for GCOS and 547 for ECOS.

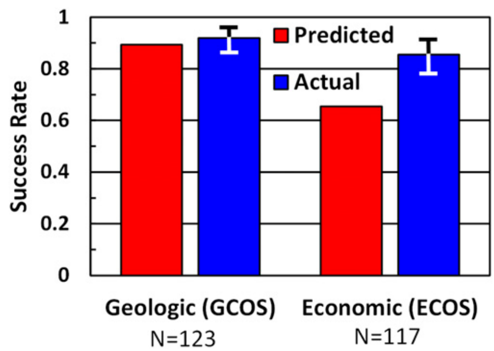

Figure 5. Success rates for unconventional wildcats (ExxonMobil Exploration Company results only; XTO Energy, Inc., wells are excluded). Insufficient data exist across risk tranches to separate the data on that basis. Overall geologic and economic success rates are higher than for conventional wildcats. It should be noted that the ultimate commerciality of some of these wildcats has greater uncertainty because of the need for extensive productivity data to establish final disposition and the high sensitivity to costs and commodity prices. The whiskers are the 95% confidence interval. ECOS = economic chance of success; GCOS = geologic chance of success; *N* = sample size.

Additionally, results to date are dominated by drilling in the United States and Canada (93 of 120 wildcats). Extrapolation of the overall results to non–North American exploration may be unreliable because of different current conditions. First, the identification of United States and Canadian unconventional plays was enabled by very large legacy well data sets, which are widely available through government agencies and vendors. For example, between 1949 and 2010, approximately 2.6 million oil and gas wells had been drilled in the United States (US Energy Information Administration, 2016). Such data are sparser or more difficult to obtain in some other countries. Second, efficient evaluation of emerging unconventional plays is dependent on a favorable regulatory and fiscal environment, ready access to drilling and completion contractors, and existing infrastructure. So although favorable geology for unconventional resource plays occurs globally, the path to commercial success may be more variable, longer, or more uncertain.

Of the 17 wildcat geologic failures, most (13) were caused by reservoir issues (generally porosity or permeability), and the remainder were caused by trap and seal issues (retention or saturation). Of these failures, 12 were unconventional play tests.

Volumetric Results

Predrill and postdrill volumetric estimates of ultimately recoverable resources have been also compared for the conventional wildcats (Figure 6). Portfolio results are affected by both the volume estimates and success rates, with the sum of the predrill risked means providing a prediction of the expected total result.

Using the sum of the risked means (entire portfolio of wells), the predrill volume estimates are somewhat biased toward the low end, 27% lower than the postdrill volume (Figure 6). If overrisking (Figure 3) is accounted for by correcting the average predrill chance of success, the error is reduced to 8%. Therefore, most of the error in prognosing cumulative volume from the sum of the risked means can be attributed to risking. As a confirmation that the error is statistically significant, a Monte Carlo simulation of the wildcat portfolio that employed the predrill ECOS and means was run (K. Steffen, 2016, personal communication). The result indicates that the actual volume result is about at a p09, a significant high-side outcome based on the predrill characterization. This supports the assumption that the predrill characterization (based on the sum of risked means) is likely statistically biased low. Note that the comparison to postdrill is not amended with future field revisions,

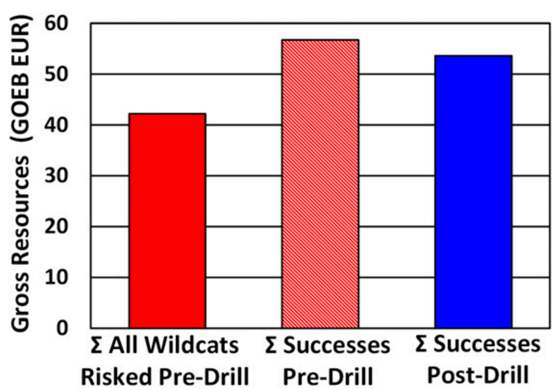

Figure 6. Comparison of predicted versus actual discovered gross recoverable volumes for 1994–2015 Exxon Mobil conventional wildcats. Two estimates for predicted volumes are employed: the sum of the risked economic means for all wildcats (left bar) and the sum of the success case economic means for the wells that were discoveries (middle bar). This latter measure is only available after drilling, but it eliminates the effect of total prospect risk. GOEB EUR = billion oil equivalent barrels estimated ultimate recovery.

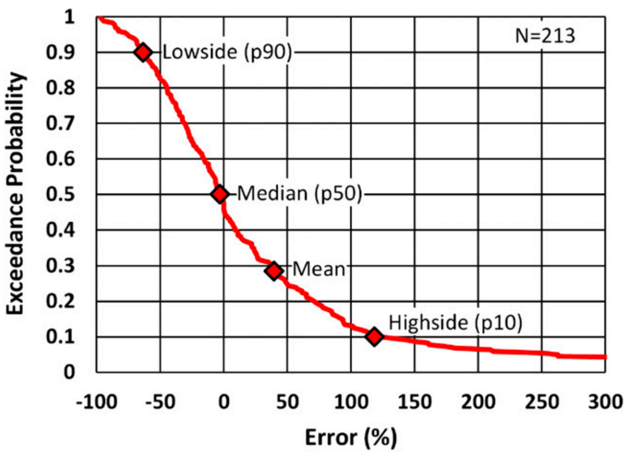

Figure 7. Cumulative frequency plot of predrill estimation error, in percent, of successful conventional wildcats. The distribution is slightly skewed, possibly indicative of a lognormal distribution. The uncertainty range is large, indicative of a high-side to low-side (10th percentile [p10]/90th percentile [p90]) ratio of approximately 7. *N* = sample size; p50 = 50th percentile.

but postdrill estimates were found to be, on average, objective relative to later estimates (described below).

As an additional test, postdrill, the sum of the predrill mean success case volumes for just the successful prospects can be compared with the sum of the postdrill volume estimates for these wells (Figure 6). The sum of the predrill mean volumes for the successful wells is 4% greater than the actual postdrill sum, a small error.

In summary, the analysis above indicates no large statistical bias on success case volume predictions. However, there is a more important effect on the overall volume prediction caused by risking.

Volumetric Uncertainty

Despite overall accurate success case volume predictions, the volume uncertainty on individual prospects is not characterized by this result. In other words, what is the typical error on individual discoveries? To evaluate this aspect of predrill assessment, the error for all successful wildcats has been calculated: (postdrill – predrill)/(predrill). This is reported as percent and plotted on a cumulative frequency distribution (Figure 7). For the entire period the volume uncertainty on individual prospects is substantial. A high-side to low-side (p10/p90) ratio of approximately 7 is indicated. To illustrate, for a 100 million bbl of oil (mean) prospect, a high-side to low-side ratio of approximately 7 could have a p90 to p10 volume range of 30 to 210 million bbl of oil.

One hundred ninety-five of the prospects had records of the predrill p10 and p90 estimates (Table 1). Ninety-four of these prospects were discoveries, and the average predrill p10/p90 was 6.6 for these wildcats. The implied p10/p90 ratio from error analysis (predrill vs. postdrill) of these same discoveries is 5.5. This data subset validates that reasonable predrill uncertainty was built into these prospects.

The predrill uncertainty range was greater for the 101 wells that turned out to be dry holes (p10/p90 = 11.7). As expected, the average predrill risk was also higher for dry holes in this subset (Table 1). The different perception of uncertainty (on average) for wells that subsequently were discoveries versus dry holes may be a secondary leading indicator of risk. Prospects with high volume uncertainty reflect more poorly understood geology because of limited understanding and data quality. In such cases, discrimination of future success and failure may also be impaired, and because most rank wildcats are dry holes, a higher failure rate would not be surprising. This was true despite a work process that explicitly separates the concepts of risk (GCOS and ECOS) and uncertainty (volume range within successful outcomes).

The overall curve is moderately skewed, with a low-probability tail toward the high-side outcomes, displacing the mean above the median (Figure 7). This is possibly consistent with the commonly held assumption that hydrocarbon volumetric distributions

Table 1. Summary of Conventional Wildcats That Had Records of Predrill Volume Range (p10–p90)

Parameter	All Wells	Dry Holes	Discoveries
Predrill p10/p90	9.3	11.7	6.6
Postdrill p10/p90	–	–	5.5
Mean predrill GCOS	0.50	0.43	0.58
Mean predrill ECOS	0.32	0.26	0.38
N (sample size)	195	101	94

For wells that were discoveries, the predrill range was similar to the implied range from postdrill analysis (difference between premean and postmean volumes). Note the estimated uncertainty was greater for wells that were subsequently dry holes. Predrill chances of success were lower for dry holes, as expected. These data are only for the wells in which a predrill range estimate was available.

Abbreviations: ECOS = economic chance of success; GCOS = geologic chance of success; p10 = 10th percentile; p90 = 90th percentile.

are approximately lognormal (Lerche, 1997; McLane et al., 2008). Note that the error function from the mean of a lognormal distribution is also lognormal in shape.

Testing the distribution on a normal quantile probability plot supports an approximately lognormal distribution. It is strongly curvilinear when values are plotted on an arithmetic axis but approximately linear when the axis is logarithmic (the postdrill/predrill ratio was tested in lieu of the error function to eliminate negative values). The best argument for volumetric assessments being lognormal is that multiplication of random distributions of variables yields a lognormal distribution (Limpert et al., 2001). Even if the parameter distributions are normal, the resultant product is approximately lognormal (Lerche, 1997). Because the hydrocarbon volumetric calculation requires the multiplication of several parameters (e.g., gross rock volume × net/gross × porosity × saturation × shrinkage × recovery efficiency) and these all have uncertainty, it is plausible to expect an approximately lognormal shape. Additionally, prospects all have a specific lower bound (the geologic minimum) but not a hard upper bound.

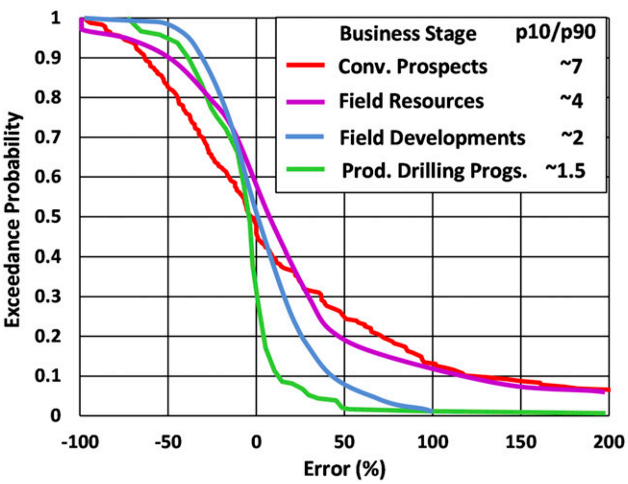

Figure 8. Cumulative frequency plot of predrill versus postdrill error for fields and prospects. Note that the uncertainty for fields is considerably less than for prospects, consistent with expectations. Also, the error or uncertainty distribution is less skewed (more symmetrical) for the three discovered field analyses. Conventional (Conv.) prospects are 1994–2015 wildcat discoveries (213 discoveries). Resource revisions are 2000–2013 technical field revisions (150 fields). Field developments are projects funded in 2005–2011 and reappraised subsequently (128 projects). Production (Prod.) drilling programs (Progs.) are conventional programs implemented in 2009–2012 (657 wells in 76 programs). p10 = 10th percentile; p90 = 90th percentile.

It should be noted that the error distribution is only slightly skewed, as evidenced by the median (p50) being near zero—a highly skewed distribution would produce a median error that would be a larger negative value. In other words, although it is sometimes invoked by explorationists, very large high-side outcomes were not common in this data.

It is instructive to compare the uncertainty estimated for prospects, with similar measures postdiscovery. Three additional databases were built that compared predicted volumes with results: total resource in fields (based on technically based revisions), field development volumes, and production drilling volumes. These showed very little error on a portfolio sum basis, similar to exploration prospects. However, all had lower uncertainty on an individual prediction basis, with high-side to low-side ratios ranging from approximately 1.5 to 4 (Figure 8). This is as expected because much more information is available postdiscovery. Additionally, the shapes of these error functions are more symmetrical, indicating more normal uncertainty distributions or at least very low skewness. On a speculative basis, this might be because extreme high-side outcomes have been eliminated in discovered fields. Of note, total field resource revisions, which include many of the discoveries in the primary data set, have a p50 near 0. This supports the earlier assertion that postdrill assessments, although subject to revision, are objective, on average.

Accurate predrill characterization of the volume uncertainty range is an important, albeit underappreciated, aspect of prospect assessment. This is because the recognition of other outcomes, especially an attainable high side, can have a major influence on the risked valuation of an opportunity. An accurate and comprehensive uncertainty analysis facilitates the development of probability weighted scenarios that yields a more complete characterization of the exploration opportunity on a technical and commercial basis. Good portrayal of the range of outcomes also enables the identification of key knowledge gaps that drives future activities and data collection (i.e., value of information, as described by Bratvold et al., 2009).

Exploration Maturity

The wildcats in this study have been classified with respect to exploration maturity in two ways: play

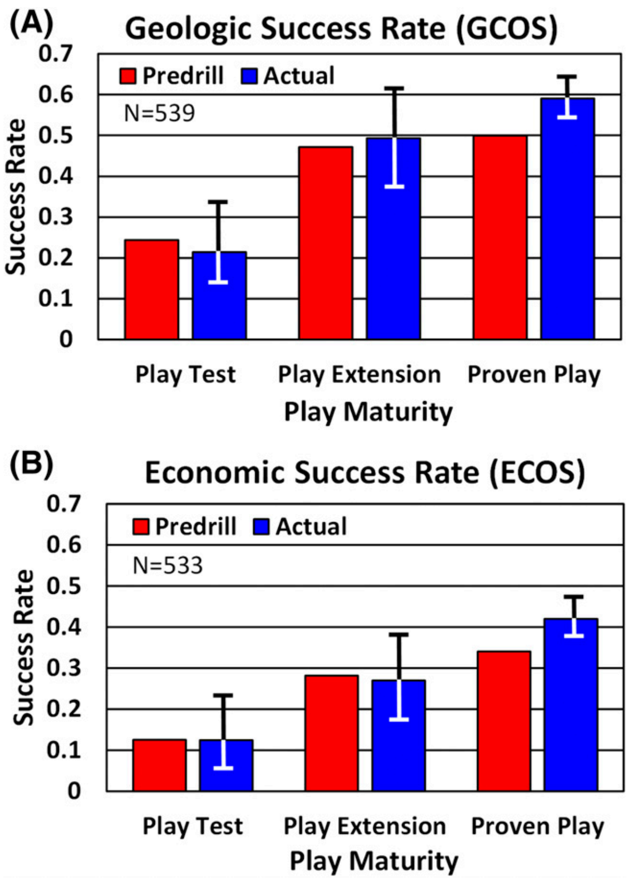

Figure 9. (A) Geologic and (B) economic success rates for conventional wildcats categorized by play maturity (play tests, play extensions, or within proven plays). Note the increase in success rate (both predicted and actual) with play maturity. The whiskers are the 95% confidence interval. ECOS = economic chance of success; GCOS = geologic chance of success; *N* = sample size.

maturity, which references petroleum system elements, and exploration stage, which is defined by the position on a cumulative discovery curve.

Regarding play maturity, wildcats are classified into three categories. Play tests are wildcats that evaluate unproven, new plays with respect to hydrocarbon system elements. Play extensions are significant step outs (generally >50 km [30 mi]) from areas where a play is established. Remaining wildcats are in proven plays, where at least geologic success is demonstrated in the vicinity. Geologic and economic success rates are strongly influenced by play maturity, increasing from play test to play extension to proven plays (Figure 9). The predrill risking was fairly accurate in aggregate across the wildcat types, although wells in proven plays were somewhat overrisked, on average.

A different criterion for exploration maturity is exploration stage, as defined by the amount of discovered hydrocarbons versus the remaining undiscovered (Rudolph et al., 2014). This can be best illustrated on a cumulative discovery volume or creaming curve for a play (Figure 10). Unproven plays are in the test stage, equivalent to play test described above. The period when the play is proven but less than approximately 50% of the total petroleum endowment is discovered is termed "emerging." The next approximately 40% is termed "core," and the final approximately 10% is termed "harvest." The typical creaming curve has a steep initial segment after a play is opened, with a flattening over time as influenced by decreasing average discovery size (Meisner and Demirmen, 1981).

Wildcats have been classified into these four exploration stages, based on local knowledge, including play assessments versus the cumulative discovered. Moreover, industry wildcats have also been classed, using lower-fidelity basin-scale criteria. Figure 11A indicates that both Exxon Mobil and industry exploration show an increase in success rates from test to emerging to core. Success rates decline in the harvest stage. This pattern is interpreted thus. With additional drilling, improving calibration and data support better success rates. However, as the inventory of quality opportunities is exhausted, more and more marginal opportunities are pursued in the harvest stage, resulting in lower success rates. Note that outpost or near-field wildcats (drilled within or very near developed fields) have been excluded from this study and are quite common in harvest-stage plays. These have a high

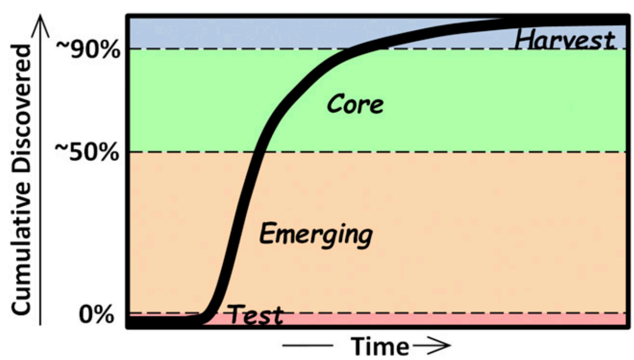

Figure 10. Idealized cumulative petroleum discovery or creaming curve with indicative exploration stages relative to the percent discovered of the ultimate endowment. Wildcats have been characterized into these four groups based on local knowledge and play assessments.

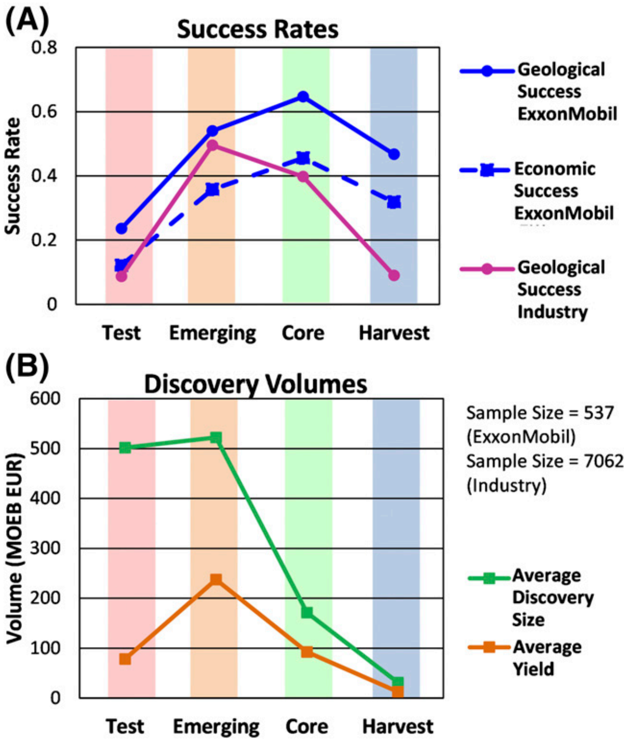

Figure 11. (A) Geologic and economic success rates by exploration stage. See Figure 10 for illustration of exploration stages. (B) Average discovery volume size and per wildcat volume yield (yield = Σ commercial volumes/Σ wildcats). Success rates increase from test to emerging to core, then decline in harvest stage. Average discovery size declines from test through harvest, with yield reaching a maximum during the emerging stage because of the combination of average field size and success rate. Industry data incorporate information from both Exxon Mobil and IHS Markit, its affiliated and subsidiary companies, and its data partners. MOEB EUR = million oil equivalent barrels estimated ultimate recovery.

success rate, but average discovery size is generally small (<10 million oil-equivalent barrels [MOEB]).

Industry success rates are somewhat lower than comparable Exxon Mobil rates, especially in the core and harvest stages. This may be related to differences in the portfolio of opportunities available to a single company versus global results. However, it is hypothesized that wildcat investment criteria across industry are variable, especially in very mature plays where established infrastructure and lower costs may enable commercialization. This is not a question of right or wrong, just an expression of different strategies; some companies may be satisfied with lower success rates and smaller discovery sizes during exploration in more mature areas, as supported by favorable economics.

Regarding volumes, there is a pattern of decreasing average commercial discovery size with exploration stage maturity. Discoveries in the test and emerging stages averaged approximately 500 MOEB recoverable per discovery versus approximately 30 MOEB in the harvest stage (Figure 11B).

Wildcat yield is an important measure of exploration efficiency. It is defined as the average commercial volume per wildcat (yield = Σ commercial volumes/Σ wildcats). Thus, the number of wildcats includes commercial discoveries, noncommercial discoveries, and dry holes (the latter two do not contribute to volumes).

Yield is maximum during the emerging stage because of large discovery size and relatively high success rates. This steadily declines through the core and harvest stages, as field size declines. These insights underline the importance of play tests, despite their low success rate—the few successful ones generate enormous value through the ensuing emerging stage. A fast follower strategy (i.e., recognizing competitors' play test successes and attempting to exploit) might be flawed in many plays because of large acreage tracts, the leveraging of improved technology to rapidly exploit a play, and the competitive advantage of early insights (e.g., deep-water Angola, Mozambique). To illustrate the importance of play tests, of the 67 discoveries that were in the emerging stage in this study, 43 (64%) were follow-up drilling to a successful play test that the company participated in.

It is also strategically important to recognize how the risk and volume profile may change during the course of exploration in a specific play. A danger exists in assuming that previous success rates and volumes are sustainable as a play matures. Very mature areas, despite deep knowledge and existing infrastructure, may be relatively unattractive unless a new concept can unlock hidden value (e.g., Johan Sverdrup discovery, Norway). Knowing when to stop to avoid value destruction is commonly a more difficult decision than when to enter a given exploration play—both local experience and the global trends described herein can be used to inform this choice.

Causes of Wildcat Failure

Based on postdrill evaluation, dry holes were categorized as having three broad causes of failure: reservoir (presence and quality), trap and seal (closure presence; top, fault, or lateral seal), and petroleum system (source presence, source maturity, migration, migration and trap timing, and hydrocarbon preservation and biodegradation). Approximately half of 246 failures can be attributed to trap and seal elements (Figure 12), dominated by fault seal and top seal, instead of closure definition. The remainder of the failures are approximately evenly split between reservoir and petroleum system elements. For conventional wildcats that failed for reservoir reasons, over 80% were related to reservoir presence (thickness, net to gross, and lithology) instead of reservoir quality (porosity or permeability). Hydrocarbon systems failures were widely distributed across source presence, maturity, migration, and timing elements.

For unsuccessful wildcats, the critical risk factors vary by play maturity. Nearly half of dry holes that were play tests or play extensions failed because of hydrocarbon system elements—source presence, maturity, migration, or timing (Figure 13). Most commonly, petroleum system risk contains a play-scale aspect, with success or failure critically influencing the outlook for adjacent prospects. In proven plays, trap and seal risk is preeminent (Figure 13). This is more commonly a prospect-specific issue. Within proven plays, reservoir and especially hydrocarbon system aspects are calibrated by regional control and therefore are somewhat lower risk. These different risk profiles may be useful to integrate into an exploration strategy. For unproven plays (play tests and play extensions), indications of a robust petroleum system (seeps and DHIs) may be of increased importance. In proven plays, trap definition (e.g., high-quality three-dimensional [3-D] seismic) might be more important. Causes of failure versus exploration stage show a similar pattern, with petroleum system dominant in the test stage and trap and seal most important in emerging, core, and harvest stages.

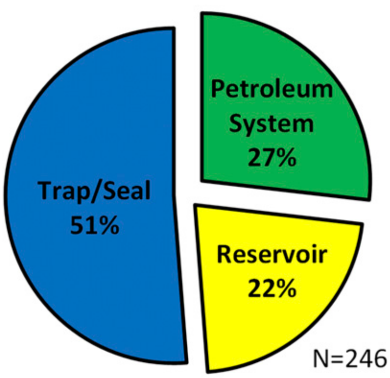

Figure 12. Causes of failure of geologic dry holes. About half of dry holes are caused by trap and seal failure. N = sample size.

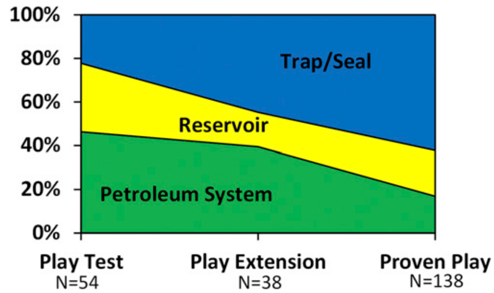

Figure 13. Reasons for wildcat failure by play maturity. Petroleum system is the most important cause of dry holes for conventional play tests and play extensions. For wildcats in proven plays, trap and seal is preeminent. N = sample size.

Technology and Geologic Factors

The correlation of success rates with other factors provides additional insights. First, prospects based on 3-D seismic data average a 10%–15% higher geologic and economic success rate (Figure 14). Success rates were also underpredicted predrill by a greater degree for wildcats with 3-D support. By itself, the application of a technology does not necessarily improve the chance of success of a specific prospect. Instead, it refines prospect characterization, including risk elements, commonly modifying the outlook to either a higher or lower chance of success. Three-dimensional seismic supports better discrimination of trap and seal risk and DHIs (where applicable), enabling the selection of lower-risk prospects within a play or venture. Many of the wildcats were associated with nondiscretionary work commitments or acreage retention; in such situations, 3-D seismic data support the identification and selection of lower-risk drilling opportunities. The underprediction of success, although good for business results, is more of a challenge because it potentially introduces some bias. Although 3-D seismic prospects were viewed as overall less risky predrill, perhaps the risking process did not take full credit for the improved confidence. For the wildcats that were unsuccessful, trap and seal was less of a factor for 3-D–supported prospects (51% vs. 62%

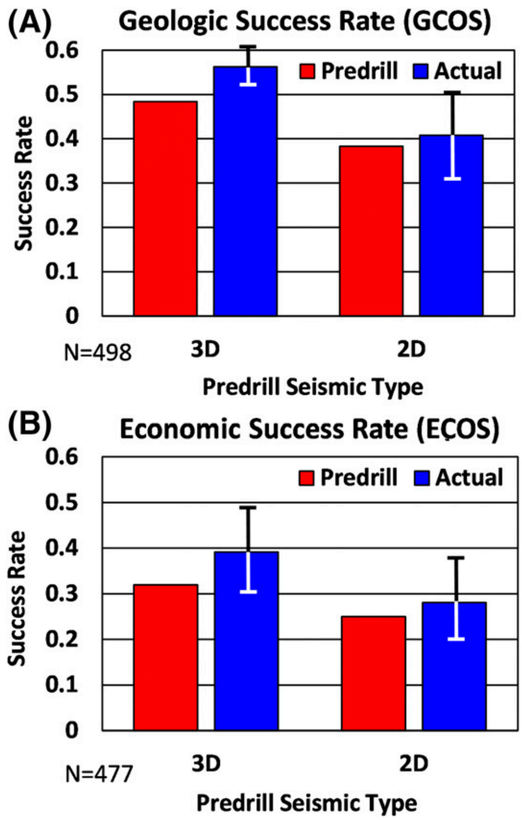

Figure 14. (A) Geologic and (B) economic success rates for conventional wildcats supported by either two-dimensional (2-D) or three-dimensional (3-D) seismic data. Three-dimensionally based wildcats experienced a 10%–15% greater success rate. The whiskers are the 95% confidence interval. ECOS = economic chance of success; GCOS = geologic chance of success; *N* = sample size.

of two-dimensional [2-D] seismic failures), possibly because of improved imaging of trap elements.

Prospects with predrill DHI support have over twice the success rate on both a geologic and economic basis (Figure 15). It should be noted that predrill DHI support was designated with a very permissive predrill criterion—marginal anomalies that were recognized predrill but designated "not valid" or "unrateable" by an internal best practice (Rudolph, 2001; Fahmy, 2006) were included. Strikingly, DHI-supported prospects were also significantly overrisked (Figure 15), with success rates some 15%–20% higher than expected. In contrast, non-DHI prospects have had a slight overprediction of success (but at the edge of statistical significance). For DHI-supported prospects that failed, very few were attributed to reservoir relative to prospects with no predrill DHI support (10% vs. 26%).

Other factors show risk differences, but these are modest and are subsidiary to the other factors described above.

- Offshore wells have 5%–10% higher average success rates. However, these prospects have a higher proportion of 3-D seismic control and DHI support, which are probably drivers.
- Geologic age (Cenozoic, Mesozoic, and Paleozoic) modestly favors Cenozoic wildcats (+5%–10% economic success rate), but these prospects much more commonly had DHI support.
- Depositional environment (deep-water siliciclastic, shallow-water to fluvial siliciclastic, and carbonate) is not correlated with meaningful differences in success rates.
- Trap types (four-way closures, three-way fault-dependent and salt-related closures, and stratigraphic

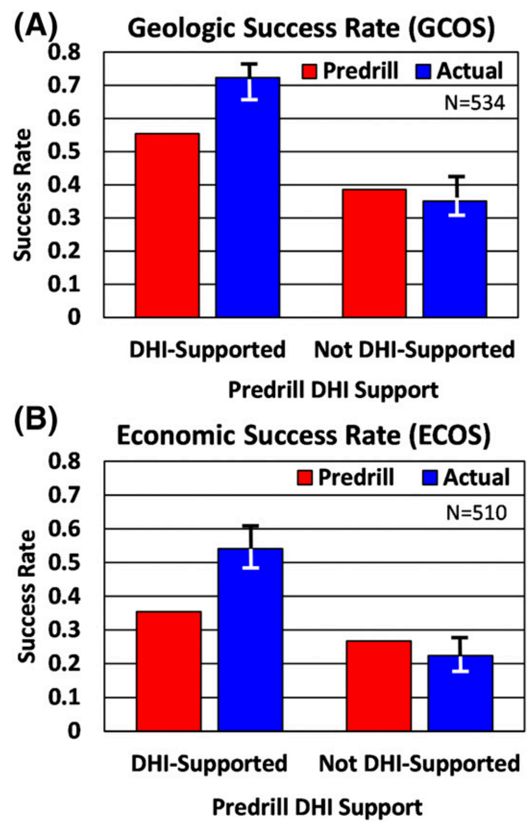

Figure 15. (A) Geologic and (B) economic success rates versus predrill direct hydrocarbon indicator (DHI) support. Both the predicted and actual success rates are significantly higher for DHI-supported prospects. Moreover, DHI-supported prospect success rates were underpredicted by a significant degree. The whiskers are the 95% confidence interval. ECOS = economic chance of success; GCOS = geologic chance of success; *N* = sample size.

closures) are not correlated with different success rates, which is somewhat surprising. However, approximately 60% of both three-way or salt-related and stratigraphic traps tested had predrill DHIs, as opposed to only approximately 30% of four-way closures. Possibly, closures judged inherently more risky were preferentially identified, matured, and drilled when they were associated with anomalies. Failure causes were not significantly different for structural versus stratigraphic traps.

- Smaller prospects (<150 MOEB) have a slightly higher success rate (+7% GCOS and +3% ECOS) than large prospects. Play tests, which average the largest prospect size but the lowest success rate, influence this result. Similarly, larger prospects that failed had a higher petroleum system causation (35% vs. 16%), which we attribute to the influence of play tests and play extensions (high petroleum system risk and largest prospects).
- Prediction of the primary commodity (oil vs. gas) was correct approximately 90% of the time, with no obvious statistical bias. This likely reflects accurate prediction of both source type and maturity, tied to well calibrations. Of greater interest, commodity results versus the primary expected petroleum source type yielded starkly different outcomes (Table 2). Type I, II, and mixed II/III sources strongly favored oil discoveries, whereas type III sources were very gas-prone. This is consistent with their expected generative nature (Van Krevelen et al., 1951). Although cracking of liquids to gas can occur at high maturity across source types (e.g., Waples, 2000), source facies seems to be a primary control for the 180 wells that have predrill source facies predictions that were audited by discovered hydrocarbons. Note that the relative number of gas versus oil outcomes does not necessarily indicate the natural occurrences of these commodities, because oil was preferentially targeted in many plays.

Table 2. Primary Commodity Discovered (Oil or Gas) versus Interpreted Predrill Source Type

Source	Oil Discoveries	Sample Size
Type I	100%	13
Type II	81%	109
Type II/III	79%	29
Type III	3%	29

Although data size is small, there is a good correspondence between the expected generative nature and the percent of the discoveries being oil.

The effect of various factors on volume uncertainty is summarized in Figure 16. Over time, there has been improvement in volumetric accuracy, with the p10/p90 ratio decreasing from approximately 9 (1994–2000) to approximately 5 (2001–2007 and 2008–2015). Although it is difficult to prove, factors that may have influenced this trend likely include improved technology (especially high-quality 3-D seismic data) and more consistent assessment practices, which were introduced over this time. Sixty-five percent of wildcats were drilled on 3-D seismic in the 1994–2000 time period; this increased to 85% for 2001–2015. There was no meaningful change in the fraction of prospects that had predrill DHI anomalies (48% vs. 47%). Note that overall success rates also increased markedly around 2002, coincident with the increased use of 3-D seismic, which may have facilitated the selection of lower-risk prospects within the portfolio (Figure 4). Improvements in seismic imaging around this time (such as prestack depth migration) may also be a factor.

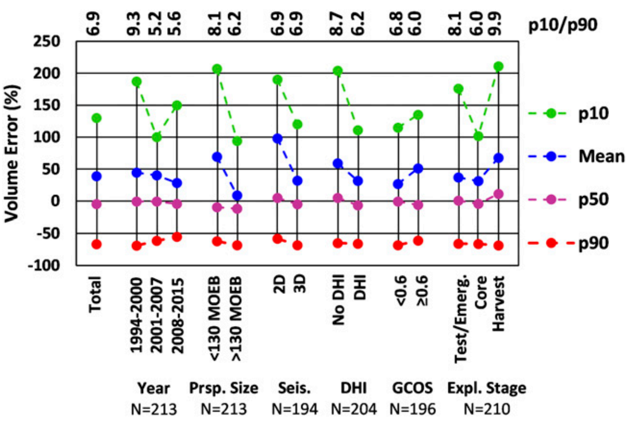

Figure 16. Conventional wildcat volume prediction uncertainty (based on predrill vs. postdrill volume errors) depicted by range plots. The total population data (on the left) are the same data described more fully in Figure 7. The implied volume uncertainty range is also illustrated for different time periods, predrill prospect (Prsp.) size, seismic (Seis.) control, direct hydrocarbon indicator (DHI) support, predrill geologic risk, and exploration (Expl.) stage. The 10th percentile/90th percentile (p10/p90) ratios are annotated at the top. 2-D = two-dimensional; 3-D = three-dimensional; Emerg. = emerging; GCOS = geologic chance of success; MOEB = million oil-equivalent barrels; *N* = sample size; p50 = 50th percentile.

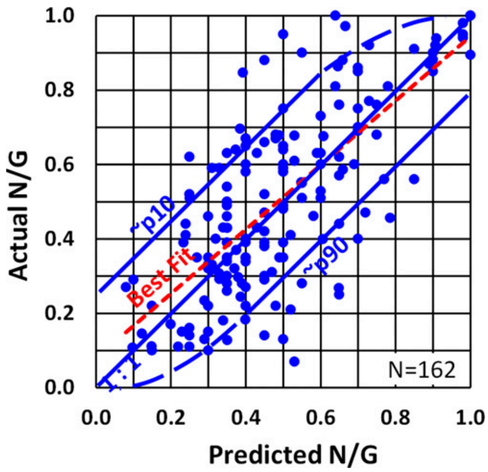

Figure 17. Predicted versus actual reservoir net-to-gross (N/G) outcomes for a subset of prospects in which these data were recovered. No statistical bias exists toward overprediction or underprediction (note linear fit vs. 1:1 correlation). However, the uncertainty is very large ($R^2 \approx 0.45$), as indicated by the implied range, which is approximately ±20%–25%. Here and in Figures 18 and 19, the 10th percentile (p10) and 90th percentile (p90) lines are approximate and are for illustrative purposes only. N = sample size.

The implied prediction uncertainty is somewhat greater for smaller prospects, which is plausible: larger leads are easier to characterize and also have the benefit of multiple reservoir targets (Figure 16). Note that the 130-MOEB criterion was chosen to split the population approximately in half. Having a predrill DHI anomaly also was correlated with improved accuracy in the predicted volumes of conventional discoveries (Figure 16). This result was consistent with the predrill evaluation of volumetric uncertainty: DHI-supported prospects had a p10/p90 ratio of approximately 6, versus 13 for non–DHI-supported prospects. Presumably, the more explicit detection of reservoir and fluids in DHI prospects leads to improved volumetric assessment. Two-dimensional versus 3-D seismic did not seem to have an effect, which is somewhat surprising. Higher-risk prospects (GCOS < 0.6) that were discoveries experienced somewhat higher volume error, implying greater uncertainty. With respect to business stage, volume uncertainty is a minimum during the core stage (Figure 16). This might be because play and prospect characteristics are well calibrated, but prospects are still material in size. During the ensuing harvest stage, although data are abundant, discoveries are quite small on average (Figure 11B), which corresponds to greater volume uncertainty.

Parameter Uncertainty

For a subset of wells, information was recovered on selected reservoir parameters used in the hydrocarbon volumetric calculation and depth prognoses (primary objective only). The parametric analysis illustrates how uncertainty can be evaluated at a more fundamental level.

Reservoir net to gross is an integral part of the predrill analysis and assessment of prospect volumes. Comparison of predrill versus actual net-to-gross outcomes for 162 wells shows that estimates were objective, neither optimistic nor pessimistic (Figure 17). However, the uncertainty is large, with ±20%–25% net to gross an appropriate estimate for the p10 to p90 range. Forty-seven predrill assessments in this data set had the input (predrill) range; the average p10 to p90 range was ±18%, not inconsistent with the postdrill results from error analysis of the larger data set.

Depth prognoses, which include both velocity and correlation or pick uncertainty, were compared with results for 141 wells (Figure 18). The average absolute error was 2.5%, with approximately 55% of the wells coming in high to prediction. The implied p10 to p90 range was ±4%–5%. An internal rule of thumb for depth uncertainty related to velocity uncertainty is ±3%, which is not wildly inconsistent with these results. Many of the greatest errors can be

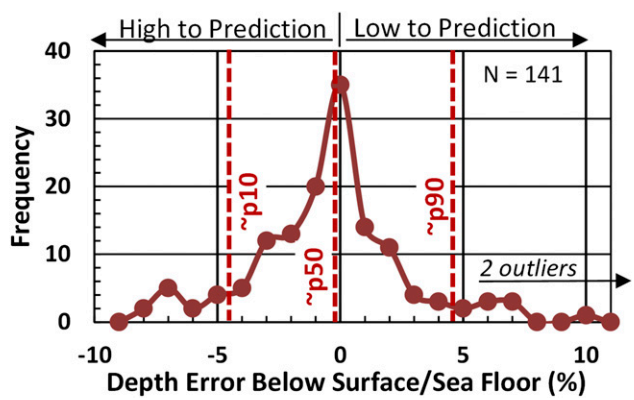

Figure 18. Frequency plot of depth prediction error, referenced to land surface (onshore) or seafloor (offshore). Points are plotted at the midpoint of 1% wide histogram bins. Wildcats that encountered their primary targets shallow to prognosis are plotted as negative values (left side of graph). Predictions are relatively objective overall but subject to significant uncertainty. Two outlying points (to the right) are included in statistics but are not plotted. N = sample size; p10 = 10th percentile; p50 = 50th percentile or median; p90 = 90th percentile.

attributed in large part to pick uncertainty: of the seven wells that exceeded 7% error, three were in areas of poor seismic imaging related to subsalt or onshore fold belts.

Average absolute error ($N = 103$) in predicted total porosity for primary reservoirs was approximately 3% porosity units, with no apparent bias (Figure 19). The implied p10 to p90 range is approximately ±4%–5%.

Average absolute error for hydrocarbon saturation ($N = 57$) was approximately 6% saturation units, with approximately 55% of the wells coming in low to expectations. The implied range is approximately ±8%. Note that this is the smallest parametric data set, and the postdrill outcomes were based on the initial log analysis (not tied to special core analysis).

CONCLUSIONS

The purpose of benchmarking is not to supplant existing evaluation processes but to trigger deeper questioning of assumptions. The oft-used financial disclaimer "past performance is not indicative of future results" certainly applies to exploration results. Moreover, statistical correlation between factors evaluated from drilling results does not prove causality.

That said, historical data on drilling results can provide initial a priori probabilities of outcomes and

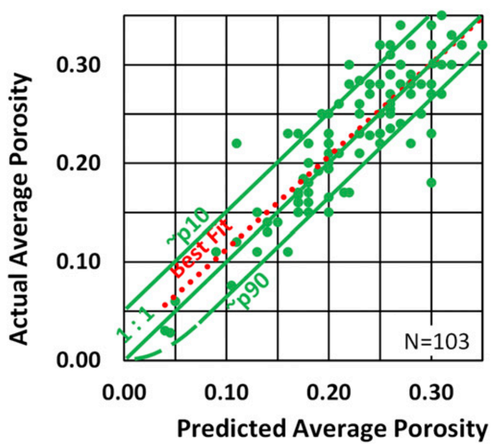

Figure 19. Predicted versus actual average reservoir porosity for a subset of conventional prospects in which these data were recovered. No strong statistical bias exists toward overprediction or underprediction (note linear fit). However, uncertainty is large ($R^2 \approx 0.75$), as indicated by the implied p10 and p90 lines, which are approximately ±4%–5%. N = sample size; p10 = 10th percentile; p90 = 90th percentile.

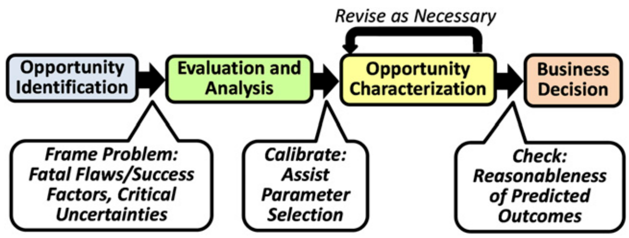

Figure 20. Use of global benchmarks in a particular opportunity evaluation, such as a prospect or new acreage opportunity.

uncertainty ranges, especially when an opportunity is further characterized as to exploration maturity, geoscience attributes, data control, etc. Wide deviation from such averages are unquestionably permissible but should be evaluated analytically. Such rules of thumb are already embedded in the use of analogs and the multitude of experience-based judgments we make on interpretations, parameters, and risk. This is merely another arrow in the quiver but one that is grounded on historical results, instead of anecdotal experience. Such prior knowledge can enhance heuristic, expert-based techniques that may be prone to cognitive biases (Baddeley et al., 2004).

At least three points exist in a geoscience analysis of a particular prospect at which benchmarks may be useful (Figure 20). Results in the play, analog play, or analog business stage can help indicate critical risk and uncertainty factors, thus helping to direct initial technical work and data collection (frame problem). Next, as base technical work nears completion, benchmarked results can assist in the selection of parameters as input into an assessment (calibrate). And finally, benchmark statistics can be powerful in the testing and possible modification of aggregate assessment and risk results (check).

Another effective use of benchmarks is for developing improved global predictive concepts and tools by quantitatively evaluating critical success or failure factors. This has the potential to unlock value via identifying new opportunities that either eluded recognition or appeared unfavorable on a risk–reward basis. Within this study, analysis of exploration drilling results validated that hydrocarbon system was the most important risk for play tests and play extensions. Although not an unexpected outcome, this helped drive recent research and development of improved technologies for recognition of petroleum systems in unproven basins or plays. Another insight was that DHI-based prospects had a significantly higher (and somewhat underestimated) success rate, even when these anomalies were of relatively poor quality. This has spawned a deliberate effort to improve the identification and characterization of subtle DHIs via improvements in technology and work practices.

To mature a prospect, one must be passionate about the idea. Unfortunately, it is easy for this ardor to cloud one's judgment and lead to overconfidence (too narrow a range of outcomes) and inappropriate optimism or pessimism (too high or low chance of success). It is our experience that the use of benchmarks, placed in a strong geologic context with specific examples, has been effective at stretching technical teams to consider alternatives and thereby develop a more objective characterization that leads to better business decisions and results.

REFERENCES CITED

Baddeley, M. C., A. Curtis, and R. Wood, 2004, An introduction to prior information derived from probabilistic judgements: Elicitation of knowledge, cognitive bias and herding, *in* A. Curtis and R. Wood, eds., Geological prior information: Informing science and engineering: Geological Society, London, Special Publications 2004, v. 239, p. 15–27, doi:10.1144/GSL.SP.2004.239.01.02.

Bratvold, R. B., J. E. Bickel, and H. P. Lohne, 2009, Value of information in the oil and gas industry: Past, present, and future: SPE Reservoir Evaluation & Engineering, v. 12, no. 4, p. 630–638, doi:10.2118/110378-PA.

Fahmy, W. A., 2006, DHI/AVO best practices methodology and applications, AAPG-SEG Distinguished Lecture, accessed February 5, 2016, http://www.seg.org/education/lectures-courses/distinguished-lecturers/fall2006/2006index.

Lerche, I., 1997, Geological risk and uncertainty in oil exploration: San Diego, California, Academic Press, 658 p.

Limpert, E., W. A. Stahel, and M. Abbt, 2001, Log-normal distributions across the sciences: Keys and clues: Bioscience, v. 51, p. 341–352, doi:10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2.

McLane, M., J. Gouveia, G. P. Citron, J. McKay, and P. R. Rose, 2008, Responsible reporting of uncertain petroleum reserves: AAPG Bulletin, v. 92, no. 10, p. 1431–1452, doi:10.1306/06040808075.

Meisner, J., and F. Demirmen, 1981, The creaming method: A Bayesian procedure to forecast future oil and gas discoveries in mature exploration provinces: Journal of the Royal Statistical Society. Series A (General), v. 144, p. 1–31, doi:10.2307/2982158.

Peel, F. J., and J. R. V. Brooks, 2016, A practical guide to the use of success versus failure statistics in the estimation of prospect risk: AAPG Bulletin, v. 100, no. 2, p. 137–150, doi:10.1306/08101515059.

Rose, P. R., 1987, Dealing with risk and uncertainty in exploration: How can we improve?: AAPG Bulletin, v. 71, no. 1, p. 1–16.

Rose, P. R., 1999, Taking the risk out of petroleum exploration-—The adoption of systematic risk analysis by oil and gas corporations during the 1990s: Leading Edge, v. 18, p. 192–199, doi:10.1190/1.1438250.

Rudolph, K. W., 2001, DHI/AVO analysis best practices: A worldwide analysis, AAPG-SEG Distinguished Lecture, accessed February 5, 2016, https://archives.aapg.org/education/dist_lect/slides/2001_02/rudolph.pdf.

Rudolph, K. W., P. J. Lehmann, and D. Erratt, 2014, Geologically complex basins: Challenges and opportunities (abs.): 21st World Petroleum Congress, Moscow, Russia, June 15–19, 2014, 11 p., accessed March 2, 2016, https://www.onepetro.org/conference-paper/WPC-21-1010.

Sykes, M. A., K. C. Hood, S. N. Salzman, and C. J. Vandewater, 2011, "Say what we mean and mean what we say": The unified upstream risk model as a force for shared understanding (abs.): AAPG International Conference and Exhibition, Milan, Italy, October 23–26, 2011, accessed February 5, 2016, http://www.searchanddiscovery.com/abstracts/html/2011/ice/abstracts/abstracts410.html.

US Energy Information Agency, 2016, Petroleum and other liquids, accessed July 22, 2016, http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=E_ERTW0_XWC0_NUS_C&f=A.

Van Krevelen, D. W., C. van Heerden, and F. J. Huntgens, 1951, Physiochemical aspects of the pyrolysis of coal and related organic compounds: Fuel, v. 30, p. 253–259.

Waples, D. W., 2000, The kinetics of in-reservoir oil destruction and gas formation: Constraints from experimental and empirical data, and from thermodynamics: Organic Geochemistry, v. 31, p. 553–575, doi:10.1016/S0146-6380(00)00023-1.

White, D. A., 1993, Geologic risking guide for prospects and plays: AAPG Bulletin, v. 77, no. 12, p. 2048–2064.