The Impacts of Supervisor Attributes and Supervision-Related Policies on Safety and Environmental Outcomes and Reporting Behavior

Christopher J. Jablonowski[a], John J. Tolle[b]

[a] Shell Exploration and Production Company, Houston, TX, U.S.A.

[b] Value Discovery LLC, Houston, TX, U.S.A.

Abstract: This paper specifies detection-controlled regression models to investigate the drivers of health, safety, and environmental (HSE) performance and reporting behavior. The analysis confirms some results from previous research and also tests new hypotheses, with emphasis on supervision-related practices and policies. Most of the results are general and thus applicable to other regions, to other operators, and very likely to other industrial sectors. The results can be used to drive decisions regarding operating practices and HSE management system policy.

Keywords: Detection-controlled Estimation, Reporting, Safety, Supervision.

1. INTRODUCTION

Health, safety, and environmental (HSE) managers are responsible for analyzing HSE performance and continuous improvement. Quantitative analysis poses special challenges because there is no theoretical basis for assumptions regarding the functional form of HSE incident phenomena, incident data is often unbalanced (few incidents), data is not collected (typically) in cases when there are no incidents, and incidents are not always reported. However, these challenges can be met with common-sense assumptions, improved data collection strategies, and advanced modeling methods.

This work specifies detection-controlled regression models similar to those employed in previous research to investigate the drivers of HSE performance and reporting behavior in oil and gas drilling [16,17]. The analysis confirms some results from previous research and also tests new hypotheses, with emphasis on supervision-related practices and policies. Most of the results presented here are general and thus applicable to other regions, to other operators, and very likely to other industrial sectors. The results can be used to drive policy decisions regarding operating practices and HSE management system policy by providing a basis to allocate resources to those policies with the largest benefit-cost ratios.

1.1. Imperfect Reporting of Safety Incidents

As described in the literature, underreporting of incidents in the workplace occurs across many sectors [26,20,25,31,27,28]. There are various reasons for underreporting, some are intentional (evasion) while others are unintentional (ignorance). Also, underreporting can occur at any level (worker, supervisor, manager). The purpose of this study is to investigate incidence and underreporting behavior in oil and gas drilling. It is acknowledged that the prospect also exists for overreporting, for example, fraudulent reports of incidents that did not occur made as an attempt to obtain a financial gain from the employer or insurance provider, but in this study it is assumed that there is no overreporting.

There is an emerging literature on the subject of incomplete detection based on the seminal work of Feinstein [10,11]. As Feinstein predicted, his model of detection controlled estimation (DCE) could be applied in various contexts. Studies have been completed in tax compliance, environmental compliance, health diagnosis, political science, and safety in oil and gas drilling [8,5,13,39,4,19,33,16,17]. The present study adds to the empirical foundation in oil and gas drilling by re-testing previous hypotheses with improved (more granular) independent variables, and by testing new hypotheses.

1.2. Implications of Imperfect Reporting

Imperfect reporting distorts the observations of incident data. A simple example demonstrates the impacts of underreporting, assuming that no fraud occurs (modified from [17]). Consider 100 hypothetical safety outcomes in Table 1. The columns represent whether or not an incident occurred, while the rows represent whether or not the incident was reported. In this unobservable "truth" case, the underreporting is evident. In practice, however, the underreported incidents are counted with the actual non-incidents. Thus, the analyst observes the data as depicted in Table 2.

Table 1. True Incident Data

		Incident Occurred?	
		Yes	No
Incident Reported?	Yes	10	--
	No	4	86

Table 2. Observed Incident Data

		Incident Occurred?	
		Yes	No
Incident Reported?	Yes	10	0
	No	0	90

Depending on the levels of imperfect reporting, the implications can be severe. The true frequency of an incident, $P(I)$, is equal to 14/100, while the analyst computes a value of 10/100. Of course the conditional probabilities are also affected. The data in Table 1 provides the *reporting rate*, defined as the conditional probability, $P(Report|Incident) = P(R|I)$. Here, this value equals 10/14, not 1 as indicated in Table 2. The complement of the reporting rate is the *underreporting rate*, $P(No\ Report|Incident) = P(NR|I)$. It is clear that in the presence of imperfect reporting, use of the data in Table 2 will distort any qualitative or quantitative analysis. If the imperfect reporting can be modeled explicitly, then more accurate assessments can be made of the true incident phenomena. Also, the analyst will learn about factors that affect the reporting rate.

2. REGRESSION MODELS OF IMPERFECT REPORTING

It is assumed that incidents are reported as the product of two independent and sequential events. First, an incident (or set of incidents) either occurs or does not occur. Second, an incident (or set of incidents) either is reported or not reported. This assumption facilitates the mathematical treatment and discussion. Two models are specified for this study.

2.1. Model of Perfect Reporting (No Underreporting, No Overreporting)

This model is specified and estimated to establish a base case for comparison, and reflects conventional practice in regression analysis of HSE incidents [11,14,15,6,34,35,22,7,21,40,18]. That is, this model estimates the case as depicted in Table 2.

Observations were collected from nine drilling rigs over a ~22 month period in 2011-2102 from one of Shell's (operator's) onshore development assets in the U.S. The unit of observation is defined as one well. Data is collected for each well i on each rig r in the study period. There are $r = 1 \ldots R$ rigs and $i = 1 \ldots N_r$ wells on each rig, $\boldsymbol{x}_{ri}$ is a $1xh$ vector of independent variables for well ri believed to affect incidence, and $\boldsymbol{\beta}$ is defined as a $hx1$ vector of coefficients to be estimated. This model specifies the incidence function as Poisson where $ln(\mu_{ri}) = \boldsymbol{x}_{ri}\boldsymbol{\beta}$. The probability for observation y_{ri} is represented as shown in Equation (1) with the resulting log-likelihood equation shown in Equation (2). Note that the marginal effect of a variable on the dependent variable, $\partial y/\partial x_h$, is equal to $\beta_h\bar{y}$.

$$Pr(Y_{ri} = y_{ri}) = \pi(y_{ri}; \mu_{ri}) = e^{-\mu_{ri}}\frac{\mu_{ri}^{y_{ri}}}{y_{ri}!} \tag{1}$$

$$L = \sum_{r=1}^{R}\sum_{i=1}^{N_r} -e^{\boldsymbol{x}_{ri}\boldsymbol{\beta}} + y_{ri}\boldsymbol{x}_{ri}\boldsymbol{\beta} - ln(y_{ri}!) \tag{2}$$

2.2. Model of Imperfect Reporting (Underreporting, No Overreporting)

This model is specified and estimated to investigate the impacts of underreporting. One set of independent variables are specified for an incidence function, while another set of independent variables are specified for a reporting function. This model requires a key assumption regarding the reporting process. When more than one incident occurs, there are three potential outcomes for reporting. One outcome is that all of the incidents are reported, a second outcome is that none of the incidents are reported, and a third outcome is that there is partial under-reporting and a subset of incidents is reported. In the derivation below, it is assumed that for each observation of the dependent variable, incidents are either all reported or all not reported, simplifying the computations.

If one allows for the possibility of partial reporting, the implications are severe. The number of conditional reporting probabilities that need to be estimated grows significantly, even when reasonable simplifying assumptions are made. In addition, the number of terms on the right hand side of the regression is in theory, infinite. For example, to compute the probability of observing one reported incident, the analyst would have to consider all potential values of incidence. The analyst could constrain this number to limit the scope of the computation, but the selection of the cutoff point would be arbitrary. For these reasons, the case of partial under-reporting is not specified here.

The incidence function is specified again as Poisson, and the reporting function is specified as a binary probit model (see [17] for a description of the probit model). $\boldsymbol{z}_{ri}$ is a $1xj$ vector of independent variables for well ri believed to affect reporting, and $\boldsymbol{\delta}$ is defined as a $jx1$ vector of coefficients to be estimated. The probability that observation y_{ri} on the dependent variable takes on a value greater than zero is shown in Equation (3) with the resulting log-likelihood function for all non-zero observations, *m*, shown in Equation (4).

$$Pr(Y_{ri} = y_{ri}) = \pi(y_{ri}; \mu_{ri})\Phi(\boldsymbol{z}_{ri}\boldsymbol{\delta}) \tag{3}$$

$$L_m = \sum_{r=1}^{R}\sum_{i=1}^{N_r} -e^{\boldsymbol{x}_{ri}\boldsymbol{\beta}} + y_{ri}\boldsymbol{x}_{ri}\boldsymbol{\beta} - ln(y_{ri}!) + ln\big(\Phi(\boldsymbol{z}_{ri}\boldsymbol{\delta})\big) \tag{4}$$

The probability that observation y_{ri} on the dependent variable takes on a value equal to zero is the sum of the probability that no incident occurred plus the probability that an incident occurred but was not reported, and this is shown in Equation (5) with the resulting log-likelihood function for all zero observations, *n-m*, shown in Equation (6).

$$Pr(Y_{ri} = y_{ri}) = \pi(y_{ri}; \mu_{ri}) + \big(1 - \pi(y_{ri}; \mu_{ri})\big)\big(1 - \Phi(\boldsymbol{z}_{ri}\boldsymbol{\delta})\big) = 1 - \Phi(\boldsymbol{z}_{ri}\boldsymbol{\delta}) + \pi(y_{ri}; \mu_{ri})\Phi(\boldsymbol{z}_{ri}\boldsymbol{\delta}) \tag{5}$$

$$L_{n-m} = \sum_{r=1}^{R} \sum_{i=1}^{N_r} ln\left(1 - \Phi(\boldsymbol{z}_{ri}\boldsymbol{\delta}) + e^{-e^{x_{ri}\boldsymbol{\beta}}} \Phi(\boldsymbol{z}_{ri}\boldsymbol{\delta})\right) \tag{6}$$

The log-likelihood for the sample is $L = L_m + L_{n-m}$ and is maximized numerically. The asymptotic covariance matrix is estimated by evaluating the negative inverse of the Hessian at the maximum likelihood estimates. The identification conditions for this family of models are derived and explored in [10].

2.3. Dependent Variable

Incident data was collected for several categories of incidents: loss of primary containment, fires, near misses, property loss and damage, unsafe acts and conditions, and injuries and illnesses. For each well *ri*, these events were summed to create the primary dependent variable, and the events were not weighted in any way. While analysis can be performed on these data individually, the authors' experience suggests that it is more appropriate to view the collection of incidents (and potential incidents) as an overall *index* of HSE performance. A second reason for aggregating the data in this way is that unaggregated data is often too unbalanced to yield reliable statistical results, that is, the dependent variable does not exhibit sufficient variability (away from 0). Figure 1 provides the distribution of the dependent variable.

2.4. Independent Variables and Hypotheses

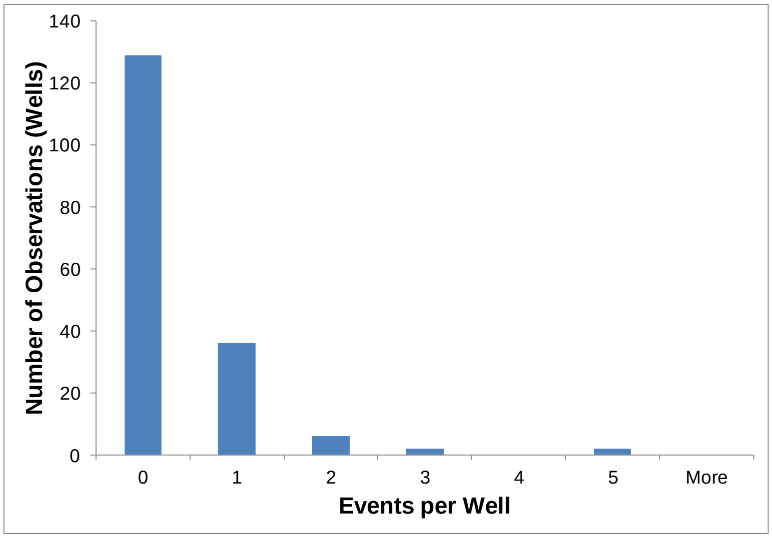
There are various taxonomies for organizing risk factors (see [1] for example). In this study, the independent variables are grouped into five categories. This structure is intended to help organize the analysis and discussion herein. For each variable, the hypothesis regarding the directional impact (sign) of the variable on incidence and reporting is stated, and whether or not the variable is expected to be statistically significant (at the 95% confidence level). Many of these expectations are based on results reported in [17], referred to in this section as "previous research."

2.4.1. Work Type

Variables in this category describe the attributes of the work being performed.

- PadSwitch: This binary variable takes a value of 1 if the well drilled on a different pad than the rig's previous well and 0 otherwise. This variable will test the hypothesis that the first well after rig-up increases the likelihood of incidents (e.g. shakedown issues). The expectation is that the sign of this variable will be positive and significant in the incidence function and insignificant in the reporting function.
- Gap: This variable is defined as the count of days between the start of the well and the end of the previous well. This variable will test the hypothesis that longer gaps between drilling operations disrupt established practices and policies and thus increase the likelihood of incidents. The expectation is that the sign of this variable will be positive and significant in the incidence function and insignificant in the reporting function. It is also recognized that there will be some correlation between this variable and the PadSwitch variable.
- WellType: This binary variable takes a value of 1 for development wells and 0 for all other wells. Previous research indicated that differences between well types in engineering design, operations, and site attributes increase the likelihood of incidents on development wells. The expectation is that the sign of this variable will be positive and significant in the incidence function and insignificant in the reporting function.
- DrillingDays: This variable is defined as the count of days from the start of the well to the end of the well. It is intended to control for exposure time and thus the expectation is that the sign of this variable will be positive and significant in the incidence function and insignificant in the reporting function.

Figure 1. Distribution of Dependent Variable



- Non-ProductiveTime: This variable is defined as the percent of total days on the well spent on non-productive activities (e.g. rig or other equipment breakdown, other unplanned events). This variable will test the hypothesis that such disruptions in normal drilling operations and switching between activities, often under increased time pressure, increase the likelihood of incidents. The expectation is that the sign of this variable will be positive and significant in the incidence function and insignificant in the reporting function.

Because the rigs were all drilling similar types of wells using similar procedures, other variables such as well design and underbalanced or managed pressure drilling could not be tested.

2.4.2. Equipment and Work Site

Variables in this category describe the rig and the location where the work is being performed.

- Rig#: Binary variables are defined for each of the nine rigs in the data set. These variables model differences in performance not captured by other variables. The expectation is that these variables will be insignificant in the incidence and reporting functions once other variables are controlled for (e.g. supervision), consistent with previous research.
- WeatherQuarter: Environmental conditions such as extreme heat or cold, or heavy rains or snows, may affect incidence. This variable is defined as a binary variable representing the calendar quarter in which the well was drilled. There are no expectations on the signs of these variables or their significance in the incident function. They are not expected to be significant in the reporting function.

The rigs were all from the same drilling contractor and were outfitted in a similar way (i.e. similar levels of automation), and the drilling sites all shared the same geography and degree of site remoteness, thus none of these variables could be tested.

2.4.3. Supervision

Variables in this category capture attributes of the onsite supervisor(s). The quality of supervision is an important factor in driving compliance with procedures [36]. How many are there? Is there well-to-well consistency? Are they owner employees or contractors? These types of questions have been investigated in several of the aforementioned studies. The results for these types of variables can be used to adjust policies on allocation of supervisory resources.

- Foreman-DaysForeman#: For each of the 59 foremen who worked within the study period, this variable is defined as the sum of days that the foreman worked on the well (“Foreman-Days”). These variables are intended to model the marginal impact of each supervisor on safety incidence. Based on previous studies, the expectation is that some supervisors will be positive (negative) and significant in the incident function, and possibly in the reporting function. An alternate version of this variable was defined as the sum of days worked on the well divided by the DrillingDays. The difference in results was negligible.
- Foreman-DaysTeam#: Some foreman worked on many of the same wells as other foreman, and of 44 of these foremen “teams” were identified. For each of these teams, this variable is defined as the sum of Foreman-Days for each foreman divided by the DrillingDays. Recent research suggests that team characteristics and communication norms can affect safety outcomes [3,24]. As is the case with the individual foreman variables, it is possible that some teams will be positive (negative) and significant in the incident function, and possibly in the reporting function.
- ForemanCount: This variable is defined as the number of different foremen who worked on the well. When more foremen are assigned to a well, it may lead to mixed messages and an increase in incidence, although previous research did not support this hypothesis. In contrast, it is also possible that having more different foremen work on a well, even if at different times, provides additional perspectives, sharing, and reinforcement of policies and best practices that decrease incidence. The expectation is that this variable will either be insignificant, or negative and significant, in the incidence function. The expectation is that for the same reasons this variable will be positive and significant in the reporting function, consistent with previous research. An interesting question is whether this supervisory *diversity* variable or the one described next that describes supervisory *concentration* will better explain incidence and reporting. Previous research did not test any hypotheses regarding concentration.
- Foreman-DaysPerRig-Day: This variable is defined as the sum of individual Foreman-Days divided by DrillingDays. For example, if Foreman#1 worked 10 days and Foreman #2 worked 15 days on a 15 day well, then this variable would equal (10+15)/15. It is intended to measure the concentration of supervision. It is possible that a larger concentration of supervision leads to better oversight and a decrease in incidence. However, it is also possible that large concentration can lead to confusion regarding who is in charge, mixed messages, and an increase in incidence. The *mixed messages* hypothesis here is different than that described in the previous variable definition because in this case the variable measures whether the messages occur at the same time. Therefore, the expectation on this variable in the incidence and reporting function is uncertain.
- ForemanConsistency: This variable is defined as the sum of individual Foreman-Days for those foremen who also worked on the previous well, divided by the sum of all Foremen-Days. Thus, a large value indicates that much of the supervision on the current well is the same as the previous well (more consistent). Previous research indicated that consistency increased the likelihood of incidents, supporting the hypothesis that the benefits of new perspectives and sharing of best practices across locations (rigs) outweighed the benefits of consistency. However, the variable as it is defined here is more refined and should produce a more reliable test of this hypothesis. Previous research used a simple binary variable to indicate whether the foremen were *exactly the same as* the previous well; for example, if only 3 out of 4 foremen were carried over from previous well, it would have been classified as not consistent. There is no expectation on the sign of this variable in the incidence function. The expectation is that the variable will be insignificant in the reporting function.

Three variables are defined to investigate the impact of foreman employment status. Previous research did not indicate any difference between operator and contractor foremen, but one hypothesis for this result was the fact that some of the contractor foreman were former operator employees and as such would reflect operator norms more so than a “pure” contractor. Foremen in each of these categories have different experiences and face different incentives, and the purpose is to test whether the degree of contractor supervision has an impact on safety incidence and/or reporting. This variable is tested in both the incidence and reporting functions, and there are no hypotheses regarding the sign of this variable.

- ForemanPureOperator: This variable is defined as the sum of Foremen-Days worked on the well by foremen who are full-time operator employees.
- ForemanPureOperatorOrFormerOperator: This variable is defined as the sum of Foremen-Days worked on the well by foremen who are full-time operator employees or a former full-time operator employees.
- ForemanPureContractor: This variable is defined as the sum of Foremen-Days worked on the well by foremen who are contractors and who have never worked for the operator.

While not a part of this study, future investigations could examine specific measures of supervisor training and competence, test for regional differences to put the spotlight on higher levels of management as suggested by others [12], and other supervisor policy options.

2.4.4. Safety Management System and Policy (SMS)

The importance of safety management systems (SMS) is self-evident. All of the rigs in this study were governed by the same SMS (e.g. inspection protocols), thus there is limited opportunity to investigate elements of the SMS. The variables in this category represent some tactics and policies deployed by the operator.

- Interventions: This variable counts the number of safety interventions made by workers and supervisors. For example, one worker may notice an unsafe act being committed by a fellow worker and intervene to stop the activity. Most companies have a mode for reporting such events, and unsafe conditions, as part of their SMS. It is commonly believed that these kinds of behaviors bolster the safety climate and improve safety performance [29], and previous research supported this hypothesis. However, it is also possible that large numbers of such interventions could be a sign of a poor safety climate. The expectation is that this variable (or percentage changes in lagged values) will be significant and negative in the incidence function. That is, larger values or percentage upticks are indicative of high levels of awareness and a good safety climate. This variable is expected to be insignificant in the reporting function. Variations of lagged specifications can also be considered as precursor candidates [32].
- WellCountOnRig: This variable is defined as the cumulative well count on each rig up to and including the current well. It is intended to capture the effect of experience on incidents and reporting. The expectation is that this variable will be negative and significant in the incidence function, and positive and significant in the reporting function. This reflects the expectation that as time passes, the operator's SMS and culture becomes more well-established, and that this improves safety and reporting performance. Previous research was inconclusive on this point. It is an important hypothesis because if indicated to be true, it may affect procurement strategy when picking up and dropping rigs.

In addition to the operator's SMS, the drilling contractor's SMS also affects safety performance and variables can be defined in a likewise fashion. As explained above, there was insufficient variability in the drilling contractor in the study period so this was not included. Future studies could include assessments of other safety policies like financial incentives for the drilling contractor [23].

2.4.5. Worker Attributes

Variables in this category provide information about the individual worker. This can be basic demographic information like age, experience in industry, and experience with the operator. The age and experience of workers are potential risk factors. One source reports that the majority of incidents involve workers with less than 5 years of experience, and that almost half of all incidents involve workers with less than 1 year on the job [2]. This category can also include measures of training and competence, and would speak to issues of training effectiveness [30,37]. Workers can also be described by levels of stress, fatigue, and workload [38]. While some of this data is collected on workers who suffer an injury, it is generally not collected for instances when no incidents occur (and this data is needed for a regression analysis). Because of this lack of data, and resource constraints which prevent its collection ex post, analysis of individual worker attributes was not included in this study.

3. REGRESSION ANALYSIS AND DISCUSSION

The model of perfect reporting was estimated first to identify *probable* drivers of incidence and/or reporting. That is, when one observes a statistically significant variable in this model, it is not discernible whether the effect is attributable to incidence or reporting behavior. However, it is a sign that the variable is probably important in one or both of the functions and careful attention is warranted in the model of imperfect reporting. When a variable does not indicate as significant in the model of perfect reporting, one cannot ignore the variable in the model of imperfect reporting. That is, it is possible that the incidence and reporting behaviors “cancel out” and thus are not observed in the model of perfect reporting.

Sample regressions for the Work Type variables are provided in Tables 3 and 4. These results are from the conventional Poisson model as specified in Equation (2); the detection-controlled models did not suggest any reporting-related effects. The results for the Work Type variables are as follows:

- The PadSwitch (Table 3) and Gap (Table 4) variables are both significant in the incidence function and have the expected signs. This result suggests that the first well after rig-up on a different pad increases the likelihood of incidents. Longer gaps between drilling operations also increase the likelihood of incidents. Because of the correlation between these variables, their effects cannot be measured precisely when both are included in the same regression.
- There is some evidence that indicates that the WellType variable is positive and significant in the incidence function, consistent with previous research. However, the effect was not consistently observed across the variety of specifications that were investigated, and in many cases it is statistically insignificant and excluded.
- DrillingDays is consistently significant (or weakly significant) as expected and is retained as a control variable for exposure time.
- Non-ProductiveTime is insignificant in the incidence function, contrary to expectations. Disruptions in normal drilling operations, switching between activities, and increased time pressure do not appear to increase the likelihood of incidents.

Table 3. Work Type Variables (a)

Variable	Coefficient Estimate	z-statistic
PadSwitch	0.6928	2.0500
WellType	1.0680	2.5200
DrillingDays	0.0430	2.4200
Non-ProductiveTime	-2.4319	-1.0000
Constant	-2.6603	-4.7000

Table 4. Work Type Variables (b)

Variable	Coefficient Estimate	z-statistic
Gap	0.0371	1.7800
WellType	0.9025	2.2000
DrillingDays	0.0524	3.0400
Non-ProductiveTime	-2.3264	-0.9500
Constant	-2.7203	-4.7000

A sample regression for the Equipment and Work Site variables is provided in Table 5. The regression includes the significant Work Type variables and two Weather variables. This result is from the conventional Poisson model as specified in Equation (2); the detection-controlled model did not suggest any reporting-related effects. The results for the Equipment and Work Site variables are as follows:

- None of the Rig# binary variables are significant in the incidence or reporting function after controlling for Foreman-Days. In regressions where Foreman-Days is excluded, some of the Rig# variables are individually significant, but the explanatory power of the models is small. When significant Foreman-Days variables are included with the Rig# variables, the Rig# variables cease to be significant and the explanatory power of the models improves ~four-fold, suggesting that the true drivers of incidents and reporting are the foremen, not the rig. Based on this result, it was concluded that in regressions that contain Rig# and no Foreman-Days variables, that the Rig# variables are merely (weak) proxies for the Foreman-Days variables because of correlations between the two variables, i.e. some foremen worked repeatedly on the same rig. The same structure and result was observed in previous research.
- There is some evidence that indicates that two of the WeatherQuarter variables are significant in the incidence function. More incidents appear to occur in the hot summer months, and fewer in the fall. While heat-related factors make sense in terms of their ability to affect HSE performance, we have no hypothesis for the apparent decrease of incidents in the fall. However, the effects was not consistently observed across the variety of specifications that were investigated, and in many cases they are both statistically insignificant and excluded.

Comparing Tables 3, 4, and 5 is instructive because they demonstrate how coefficient estimates and significance tests change when alternate specifications are estimated. This is typical for all regression-based analysis, and it puts extra demands on the analyst to explain inconsistent and/or awkward results. In this case, the results are consistent across specifications.

Table 5. Equipment and Work Site Variables

Variable	Coefficient Estimate	z-statistic
Gap	0.0454	2.2000
WellType	1.1884	2.8300
DrillingDays	0.0688	3.6200
3Q (summer)	0.7101	2.4100
4Q (fall)	-0.8378	-1.9800
Constant	-3.5131	-5.3500

The results for the Supervision and SMS variables are based on the detection-controlled model as specified in Equations (4) and (6) because the analysis suggests some reporting-related effects for these variables. In most cases the regressions include the previously identified significant Work Type and Equipment and Work Site variables, although the coefficients are not measured precisely in all regressions, especially for regressions with larger numbers of independent variables. In some cases this is explained by correlations between variables, otherwise, we attribute the fluctuations between specifications to the complexity of the incident- and report-generating processes (e.g. excluded variables, joint effects, etc.), and to the numerical complexity of the detection-controlled model (for example, in some specifications the model does not converge). The results for the Supervision variables are as follows:

- Several of the individual foremen (Foreman-DaysForeman#) are significant in the incidence function, supporting the hypothesis that some foremen have a unique impact on safety performance. There is some evidence to support differential reporting behavior as shown in Table 6. In this example result, Foreman#12 appears to have more incidents relative to other foremen, and also appears to be less likely to report an incident once it occurs. But the results for individual foremen are somewhat sensitive to the specification, i.e. the choice and number of coefficients being estimated. Also, many foremen have small

Foreman-Days totals, making it difficult to estimate their individual impacts with precision. Interpreted in their totality, the results do not suggest significant differential reporting behavior between foremen. Both results are consistent with previous research.

Table 6. Foremen Variables

Variable	Coefficient Estimate	z-statistic
Incidence Function		
Gap	0.0247	0.9500
WellType	0.8643	1.9100
DrillingDays	0.0673	3.1800
3Q (summer)	0.4326	1.3300
4Q (fall)	-0.8528	-1.9100
Foreman#12	0.3268	4.8000
Foreman#43	0.0601	1.9200
Constant	-3.1622	-4.4600
Reporting Function		
Foreman#12	-0.3777	-2.5400
Foreman#35	-0.1488	-0.9400
Constant	1.8768	2.2300

- The foreman team variables (Foreman-DaysTeam#) were defined to test hypotheses about team impacts on safety performance. In some cases, the foreman variables and the team variables are correlated, and additional analysis was required to understand the structure of the relationship. Based on this analysis, it is clear that significant foreman teams (in incidence and/or reporting) are driven by the significance of their constituents; i.e. all but one significant teams have at least one individually significant constituent. In the one case where a significant team was comprised of two individually insignificant constituents, there is an apparent synergy between the two foremen that improved their joint performance. In summary, there does not appear to be a strong team impact on incidence or reporting, rather, it is the individuals who drive these outcomes. These variables were retained and used in cases where individual foremen variables were highly correlated.
- The evidence for the ForemanCount variable indicates that when more foremen work on a well there is a decrease in the likelihood of incidents and an increase in reporting, consistent with expectations. This result indicates that when additional foremen work on a well, the diversity of supervision serves to reinforce policies and best practices rather than to introduce uncertainty from mixed messages.
- The Foreman-DaysPerRig-Day variable was insignificant in the incidence and reporting functions. This result indicates that larger supervisory concentration does not yield a measurable impact, positive or negative, on incidence or reporting. It is important to note however that this variable is somewhat tightly clustered around a value of 2, potentially affecting the precision of the estimate. Also, when the value falls below this value it is not by much, therefore no conclusions can be drawn about the impact of concentrations less than 2.
- As described above, previous research indicated that foreman consistency increased the likelihood of incidents. This was a somewhat controversial result and it was desired to revisit this question in the present study. The previous research used a simple binary variable to indicate whether the foremen were *exactly the same as* the previous well. The ForemanConsistency variable defined above is more refined and provides a more definitive test of this hypothesis. The results using the new variable definition indicate that consistency is significant in the incidence function, and that more well-to-well consistency decreases the likelihood of incidents. There is no impact on reporting.

- The analysis of the three variables that describe foreman employment status indicate that employment status does not significantly affect incidence or reporting. This result was consistent whether individual foreman variables were included or not included in the regressions, reducing the risk that correlations between the two sets of variables was affecting the result.
- The Interventions variable is significant and positive in the incidence function, contrary to expectations. This result suggests that higher levels of intervention can be interpreted here as an indicator of a breakdown in the safety climate. By itself, this information is not very useful because each rig develops different norms regarding the level of reporting. Percentage changes in Interventions was also investigated but were not found to be significant. The Interventions variable is not significant in the reporting function.
- The WellCountOnRig variable was defined to test the hypothesis that as additional wells are drilled, the operator's SMS and culture becomes more well-established, and that this improves safety and reporting performance. The variable is significant and negative in the incidence function, consistent with expectations. There is weak evidence that the variable is positive and significant in the reporting function.

An auxiliary regression was specified to investigate the relationship between Interventions and the WellCountOnRig variable. A strong negative relationship was discovered; Interventions decrease the longer a rig is in the fleet. A typical comprehensive regression is shown in Table 7.

Table 7. Comprehensive Regression (example)

Variable	Coefficient Estimate	z-statistic
Incidence Function		
Gap	0.0554	1.9900
DrillingDays	0.0397	1.4700
Foreman#6	0.1016	1.8800
Foreman#35	0.2761	2.1700
Foreman#43	0.1461	2.7100
Foreman#45	0.1432	2.9000
Foreman#55	0.1940	3.2000
Foreman#56	-0.0610	-1.2500
ForemanTeam7/16	-2.8114	-1.4500
ForemanCount	-0.4660	-2.1500
ForemanConsistency	-1.4864	-1.7800
WellCountOnRig	-0.1037	-2.8600
Constant	1.2499	1.1000
Reporting Function		
ForemanCount	2.9462	2.0000
Foreman#12	0.7016	0.8800
Foreman#35	-0.3663	-1.3800
WellCountOnRig	0.2174	1.2200
Constant	-10.9078	-1.9900

Table 8 contains a summary of the expectations and findings for each independent variable. The first column lists the variable name, the second column summarizes the expectations, and the third column summarizes the findings.

Table 8. Summary of Individual Variable Expectations and Results

Variable	Incidence Function EXPECTATION	Incidence Function RESULT
PadSwitch	+	+
Gap	+	+
WellType (dev=1)	+	+ (weak)
DrillingDays	+	+
Non-ProductiveTime	+	insignificant
Rig#	insignificant	insignificant
WeatherQuarter	?	mixed (weak)
Foreman-DaysForeman#	mixed	mixed
Foreman-DaysTeam#	mixed	insignificant
ForemanCount	insignificant or -	-
Foreman-DaysPerRig-Day	?	insignificant
ForemanConsistency	?	-
ForemanPureOperator	?	insignificant
ForemanPureOperatorOrFormerOperator	?	insignificant
ForemanPureContractor	?	insignificant
Interventions	-	+
WellCountOnRig	-	-
Variable	**Reporting Function EXPECTATION**	**Reporting Function RESULT**
PadSwitch	Insignificant	insignificant
Gap	insignificant	insignificant
WellType	insignificant	insignificant
DrillingDays	insignificant	insignificant
Non-ProductiveTime	insignificant	insignificant
Rig#	insignificant	insignificant
Weather	insignificant	insignificant
Foreman-DaysForeman#	mixed	mixed (weak)
Foreman-DaysTeam#	mixed	insignificant
ForemanCount	+	+
Foreman-DaysPerRig-Day	?	insignificant
ForemanConsistency	insignificant	insignificant
ForemanPureOperator	?	insignificant
ForemanPureOperatorOrFormerOperator	?	insignificant
ForemanPureContractor	?	insignificant
Interventions	Insignificant	insignificant
WellCountOnRig	+	+ (weak)

As reported in Table 8, there are only a few variables that affect reporting. Using the regression model reported in Table 7, it is possible to compute the probability of a false negative for each zero observation, *P(Incident|No Report)*, or *P(I|NR)*, using the notation from the Introduction. The result provides a general indication of whether imperfect reporting is a significant problem. The probability is defined in Equation (7) using the notation from the Introduction

$$
\begin{aligned}
P(I|NR) &= \frac{P(I)P(NR|I)}{P(NR)} = \frac{\left(1-\pi(y_{ri};\mu_{ri})\right)\left(1-\Phi(\boldsymbol{z}_{ri}\boldsymbol{\delta})\right)}{\pi(y_{ri};\mu_{ri})+\left(1-\pi(y_{ri};\mu_{ri})\right)\left(1-\Phi(\boldsymbol{z}_{ri}\boldsymbol{\delta})\right)} \\
&= 1-\frac{\pi(y_{ri};\mu_{ri})}{1-\Phi(\boldsymbol{z}_{ri}\boldsymbol{\delta})+\pi(y_{ri};\mu_{ri})\Phi(\boldsymbol{z}_{ri}\boldsymbol{\delta})}
\end{aligned}
\tag{7}
$$

The average probability for all zero observations is 7%, suggesting that imperfect reporting is not a significant problem in this asset. This result suggests that future analysis probably can be completed without the more complex detection-controlled models without introducing significant bias. However, for definitive results, the detection-controlled estimates are always recommended.

4. CONCLUSION AND RECOMMENDATIONS

The results of this analysis are largely consistent with previous research, strengthening the case for action on specific points [17]. Some of the results are specific to the operator or the asset, but most are general and thus applicable to other regions, to other operators, and very likely to other industrial sectors. Based on these results, the following actions are recommended for all drilling operations:

- Refresh the focus on safety after all rig moves between drilling pads and extended delays between wells by organizing a formal HSE engagement event at the rig site before the next well starts.
- Identify specific differences in exploration and development wells that have the potential to cause differences in incident rates, and engage the engineering and operations staff to mitigate these risks.
- Institute a refresher course for all foremen and the workforce prior to the summer season to emphasize heat-related hazards.
- Engage/interview foremen who were identified as more or less likely to have incidents to ascertain the potential drivers of these performance differences.
- Assign additional foremen to each well to provide additional perspectives, sharing, and reinforcement of policies and best practices. Currently, about four foremen work on any one well. Increasing the number of different foremen who work on a well can be accomplished by splitting hitches between wells, or by having more foremen on location at the same time.
- Maintain some supervisory consistency on each rig by assigning one or more of the foremen on the previous well to the current well. In cases where this is not possible, organize a formal HSE engagement event at the rig site before the next well starts.
- Continue the current contractor HSE on-boarding process which appears to be successful in ensuring equivalent incident and reporting performance with the operator’s full-time staff.
- Retain rigs for longer terms to firmly establish the SMS and reporting norms, i.e. a “first in, last out” model.

A final note of caution is needed regarding the use of this kind of information. First, when a relationship between incidents and a variable is identified and an intervention plan or policy is enacted to reduce risk, then over time the relationship between incidents and the variable will degrade and ultimately be eliminated if the intervention plan or policy is effective. For example, if it is recognized that switching pads increases the likelihood of incidents, and a new policy is effectively implemented to refresh the safety focus in such cases, then switching pads will not be identified as a risk if the analysis is repeated in the future. But this should not be interpreted as evidence that switching pads no longer increases the likelihood of incidents, rather it should be interpreted as evidence that the intervention policy is working. Also, this same phenomenon makes it difficult to identify risk factors that are already being mitigated by some policy. In this case, the lack of statistical evidence would not be a sufficient reason to alter or cancel an existing mitigation policy that is otherwise believed to be working.

5. ACKNOWLEDGEMENTS

The authors thank our colleagues who contributed technical expertise and data: Cody Buyer, Eryn Clark, Jonathan Dallaire, Cesar Gongora, James Goodwyne, Kevin Hoffman, Lake Johnson, Gregory Knott, R.J. Mendoza, Jose Mota, Brendan O'Shea, Donald Patton, Bobby Ramos, and Linda Randle. All conclusions, errors, and omissions remain the sole responsibility of the authors.

6. REFERENCES

[1] Altabbakh, Hanan, Murray, Susan, Grantham, Katie, Damle, Siddharth. 2013. Variations in Risk Management Models: A Comparative Study of the Space Shuttle Challenger Disaster. Engineering Management Journal, 25 (2):13-24.

[2] Are We Losing Ground in Rig Safety? Some Operators Say 'Yes." 1998. *Drilling Contractor*, May/June: 18.

[3] Azarkhil, Mandana, Mosleh, Ali. 2012. Modeling and Simulation of the Impact of Team Characteristics on Crew Performance. Probabilistic Safety Assessment and Management Conference, Helsinki, Finland, 25-29 June 2012.

[4] Bradford, W.D., Kleit, A.N., Krousel-Wood, M.A., Re, R.N. 2001. Testing Efficacy with Detection Controlled Estimation: An Application to Telemedicine. *Health Economics*, 10: 553-564.

[5] Brehm, J., Hamilton, J.T. 1996. Noncompliance in Environmental Reporting: Are Violators Ignorant, or Evasive, of the Law?" *American Journal of Political Science*, 40 (2): 444-477.

[6] Chunlin, H., Chengyu, F. 1999. Evaluating Effects of Culture and Language on Safety. *J. Pet Tech*, April.

[7] Conchie, S., Donald, I. 2006. The Role of Distrust in Offshore Safety Performance. *Risk Analysis*, 26 (5): 1151-1159.

[8] Erard, B. 1997. Self-selection with Measurement Errors: A Microeconometric Analysis of the Decision to Seek Tax Assistance and Its Implications for Tax Compliance. *Journal of Econometrics*, 81: 319-356.

[9] Feinstein, J. 1989. The Safety Regulation of U.S. Nuclear Power Plants: Violations, Inspections, and Abnormal Occurrences. *Journal of Political Economy*, 97 (1): 115-154.

[10] Feinstein, J. 1990. Detection Controlled Estimation. *Journal of Law and Economics*, 33: 233-276.

[11] Fleming, M., Flin, R., Mearns, K., Gordon, R. 1996. The Offshore Supervisor's Role in Safety Management: Law Enforcer or Risk Manager. Paper SPE 35906 presented at the Third International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production, New Orleans, LA, USA, 9-12 June.

[12] Fruhen, L.S., Mearns, K.J., Kirwan, B., Flin, R. 2012. Skills and Traits as Contributors to Senior Managerial Safety Commitment. Probabilistic Safety Assessment and Management Conference, Helsinki, Finland, 25-29 June 2012.

[13] Helland, E. 1998. The Enforcement of Pollution Control Laws: Inspections, Violations, and Self-Reporting. *The Review of Economics and Statistics*, 80 (1): 141-153.

[14] Iledare, O., Pulsipher, A., Dismukes, D., Mesyanzhinov, D. 1997. Oil Spills, Workplace Safety and Firm Size: Evidence from the U.S. Gulf of Mexico OCS. *The Energy Journal*, 18 (4): 73-89.

[15] Iledare, O., Pulsipher, A., Dismukes, D., Mesyanzhinov, D. 1998. Safety and Environmental Performance Measures in Offshore E&P Operations: Empirical Indicators for Benchmarking. Paper SPE 49153 presented at the SPE Annual Technical Conference and Exhibition, New Orleans, LA, USA, 27-30 September.

[16] Jablonowski, C. 2007. Employing Detection Controlled Models in Health and Environmental Risk Assessment: A Case in Offshore Oil Drilling. *Journal of Human and Ecological Risk Assessment*, 13 (5): 986-1013.
[17] Jablonowski, C. 2011(a). Statistical Analysis of HSE Incidents and the Implications of Imperfect Reporting. *SPE Drilling & Completions*, 26 (2): 278-286.
[18] Jablonowski, C. 2011(b). Using Regression Analysis to Relate Safety and Environmental Outcomes to Incidence Factors. *SPE Projects, Facilities & Construction*, March: 33-40.
[19] Kleit, A.N., Ruiz, J.F. 2003. False Positive Mammograms and Detection Controlled Estimation. *Health Services Research*, 38 (4):1207-1228.
[20] Leigh, J.P., Marcin, J.P., Miller, T.R. 2004. An Estimate of the U.S. Government's Undercount of Nonfatal Occupational Injuries (Review). *Journal of Occupational and Environmental Medicine*, 46 (1): 10-18.
[21] Malallah, S. 2009. Leadership Influence in Safety Change Effort. Paper IPTC 13816 presented at the International Petroleum Technology Conference, Doha, Qatar, 7-9 December.
[22] Mearns, K., Whitaker, S., Flin, R. 2001. Benchmarking Safety Climate in Hazardous Environments: A Longitudinal, Inter-organizational Approach. *Risk Analysis*, 21 (4): 771-786.
[23] Osmundsen, Petter, Toft, Anders, Dragvik, Kjell Agnar. 2006. Design of Drilling Contracts-Economic Incentives and Safety Issues. Energy Policy, 34: 2324-2329.
[24] Park, Jinkyun, Jung, Wondea, Yang, Joon-Eon. 2012. The Use of SNA Metrics to Investigate the Relationship Between the Characterisitcs of Crew Communications with the Associated Crew Performance. Probabilistic Safety Assessment and Management Conference, Helsinki, Finland, 25-29 June 2012.
[25] Phimister, J.R., Bier, V.M., Kunreuther, H.C. (eds.). 2004. Accident Precursor Analysis and Management: Reducing Technological Risk Through Diligence. Washington, D.C.: The National Academies Press.
[26] Pransky, Glenn, Snyder, Terry, Dembe, Allard, Himmelstein, Jay. 1999. Under-Reporting of Work-Related Disorders in the Workplace: A Case Study and Review of the Literature. Ergonomics, 42 (1): 171-182.
[27] Probst, T.M., Brubaker, T.L., Barsotti, A. 2008. Organizational Injury Rate Underreporting: The Moderating Effect of Organizational Safety Climate. *Journal of Applied Psychology*, 93 (5): 1147-1154.
[28] Probst, Tahira M., Estrada, Armando X. 2010. Accident Under-Reporting Among Employees: Testing the Moderating Influence of Psychological Safety Climate and Supervisor Engorcement of Safety Practices. *Accident Analysis and Prevention*, 42: 1438-1444.
[29] Pullwitt, Tanja, Mearns, Kathryn. 2008. Safety Interventions-A Status Challenge for Managers and Supervisors. Paper SPE 111801 presented at the SPE International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production, Nice, France, 15-17 April.
[30] Ramirez, Alberto E., Prasad, Dharmesh. 2000. Safety Training Effectiveness. Paper SPE 61251 presented at the SPE International Conference on Health, Safety, and the Environment in Oil and Gas Exploration and Production, Stavanger, Norway, 26-28 June.
[31] Rosenman, K.D., Kalush, A., Reilly, M.J., Gardiner, J.C., Reeves, M., Luo, Z. 2006. How Much Work-Related Injury and Illness is Missed by the Current National Surveillance System? *Journal of Occupational and Environmental Medicine*, 48 (4):357-365.
[32] Saltmarsh, Elizabeth A., Saleh, Joseph H., Mavris, Dimitri N. 2012. Accident Precursors: Critical Review, Conceptual Framework, and Failure Mechanisms. Probabilistic Safety Assessment and Management Conference, Helsinki, Finland, 25-29 June 2012.
[33] Scholz, J.T., Wang, C.L. 2006. Cooptation or Transformation? Local Policy Networks and Federal Regulatory Enforcement. *American Journal of Political Science*, 50 (1): 81-97.
[34] Shultz, J. 1999. The Risk of Accidents and Spills at Offshore Production Platforms: A Statistical Analysis of Risk Factors and the Development of Predictive Models. Doctoral Dissertation, Carnegie Mellon University.
[35] Schultz, J., Fischbeck, P. 1999. Predicting Risks Associated with Offshore Production Facilities: Neural Network, Statistical, and Expert Opinion Models. Paper SPE 52677 presented at the SPE/EPA Exploration and Production Environmental Conference, Austin, TX, USA, 28 February-3 March.
[36] Skaugrud, Ida, Dahl, Oyvind, Olsen, Espen. 2012. A Qualitative Study of Organizational Factors Influencing Compliance with Procedures. Probabilistic Safety Assessment and Management Conference, Helsinki, Finland, 25-29 June 2012.

[37] Skjerve, Ann Britt, Bye, Andreas. 2012. Towards Guidance for Assessment of “Training” as a Performance-Shaping Factor in Human Reliability Analysis. Probabilistic Safety Assessment and Management Conference, Helsinki, Finland, 25-29 June 2012.
[38] Sneddon, A., Mearns, K., Flin, R. 2006. Safety and Situation Awareness: “Keeping the Bubble” in Offshore Drilling Crews. Paper SPE 98629 presented at the International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production, Abu Dhabi, U.A.E., April 2-4.
[39] Stafford, Sarah L. 2003. Assessing the Effectiveness of State Regulation and Enforcement of Hazardous Waste. *Journal of Regulatory Economics*, 23 (1): 27-41.
[40] Winter, J., Owen, K., Read, B., Ritchie, R. 2010. How Effective Leadership Practices Deliver Safety Performance And Operational Excellence. Paper SPE 129035 presented at the SPE Oil and Gas India Conference and Exhibition, Mumbai, India, 20-22 January.