A Report to the Secretary of Transportation

Pipeline Integrity Management

An Evaluation to Help Improve PHMSA's Oversight of Performance-Based Pipeline Safety Programs

Rick Kowalewski
October 31, 2013

Contents

Executive summary

In the aftermath of the natural gas pipeline explosion and fire in San Bruno, California on September 9, 2010, the National Transportation Safety Board (NTSB) identified several concerns with the Pipeline and Hazardous Materials Safety Administration's (PHMSA's) oversight of performance-based safety programs, and recommended that the Secretary of Transportation conduct an audit to address these concerns.[1] The most extensive performance-based program for pipeline safety today is focused on *integrity management*.

San Bruno, CA neighborhood after the 2010 pipeline accident

Integrity management (IM) is a performance-based, process-oriented program for managing the safety and environmental risks associated with oil/hazardous liquid and natural gas pipelines in the U.S. The IM rules specify how pipeline operators must identify, prioritize, assess, evaluate, repair and validate the integrity of their pipelines that could—in the event of a leak or failure—affect High Consequence Areas (HCAs) within the United States. HCAs include population areas; and (for hazardous liquid pipelines) areas that are unusually sensitive to environmental damage or commercially navigable waterways. Hazardous liquid pipelines have been subject to the IM rules since 2001,[2] and gas transmission pipelines have been subject to the rules since 2004.[3] This evaluation does not address gas distribution systems, which were not subject to IM rules until 2011.

Pipelines have been a comparatively safe mode of transportation over the last several decades, with relatively few deaths and injuries while transporting extremely large quantities of energy products across the country. Nevertheless, they present a substantial threat of low-probability, high-consequence accidents with very high public concern when these accidents occur. The IM program was developed by PHMSA's Office of Pipeline Safety (OPS) and intended to target especially these kinds of risks.

The scope, objectives, and methodology of the evaluation

This evaluation was planned initially in 2006[4] as part of a normal process of continuous improvement in the program. However, after the San Bruno incident, the scope and objectives have been targeted more specifically to respond to NTSB's more recent recommendations.

The objective of this evaluation is to provide a sound, objective basis for improving PHMSA's inspection and oversight of pipeline operators' integrity management (IM) programs, including:

- expanding the use of meaningful metrics,
- ensuring the completeness and accuracy of information used in operators' IM programs,
- incorporating operators' leak, failure, and incident data in evaluating their risk models,
- setting goals for pipeline operators and tracking performance against those goals, and
- improving inspection protocols for integrity management inspections.

The evaluation examined how well the current program is *designed* to achieve its objectives, the extent to which it is *implemented* as designed, *assumptions* and *external factors* that might influence program outcomes, *unintended effects* that might result from the operation of the program, and evidence that could demonstrate the program's contribution to the *safety outcomes* that are occurring. It also examined PHMSA's enforcement policies and procedures, and the standard of review for compliance with performance-based regulations.

I reviewed legislation, rulemaking, past evaluations, investigation reports, and program documentation to ground the evaluation of program development, design, and implementation. I analyzed pipeline safety accident/incident data, annual reports from pipeline operators, inspection results and enforcement data to help quantify safety outcomes and program activities. I developed a program logic model (on page 20) to help evaluate the program design and assumptions.

I interviewed inspectors, managers, and technical staff from PHMSA as well as safety professionals from several pipeline companies, to help clarify how the program actually works, and to get their perspectives on strengths and weaknesses in the program. And I examined some of the risk assessment literature, external standards, and other agencies' practices—all to develop the factual basis for understanding the current program and ideas for improving it.

Overview of the evaluation results

In many ways, integrity management was a bold move into a new way of managing pipeline safety. The IM program is based on fundamentally sound logic—that companies should be responsible for managing their own risks, with regulatory oversight of processes, systems and performance. But experience with different performance-based regulatory (PBR) approaches has revealed a number of common challenges that can limit their effectiveness. Many agencies have struggled to make PBR work. It was in this context that the IM program was launched.

With IM, performance-oriented standards were added to the body of pipeline safety regulations, bringing new requirements to assess and repair the physical infrastructure, evaluate risks, take actions to reduce those risks, and put in place systems for continuous improvement. The IM rules aimed to put the burden squarely on the operator to manage its risks, and provided flexibility to do so.

Nearly everyone[5] associated with the IM program believes it is effective.[6] They point particularly to the thousands of anomalies and defects that have been found and fixed as a result of the program—commonly viewed as "*accidents avoided*"—and to improvements in technology that have been spurred by IM.

At the same time, *we are not seeing clear evidence of the safety outcomes the program expected* when the IM rules were published—particularly for gas transmission pipelines. On these systems, some indicators suggest that safety risk might be rising instead of declining—particularly high consequence accidents and material failures that were a special focus of the rules. Over 27 years of data, 18/20 (90%) of gas transmission incidents with >$10 million damage occurred in eight years *after* IM was in effect, after adjusting for inflation.

Accidents/ Incidents & Consequences	Hazardous Liquid Pipelines	Gas Transmission Pipelines
Accidents/ Incidents	~	✗
High Conseq. Incidents	~	✗
Deaths	~	~
Injuries	~	~
Property Damage	~	✗
Corrosion Failure	~	~
Material Failure	~	✗

Trends before/after IM implementation

Across the gas and liquid sectors, the six highest consequence accidents on record have been *since* IM was put in place.

Where the trends seem to be positive, they are much better explained by the damage prevention program[7] than by integrity management. The analysis examined many external factors (more pipe, aging pipe, newer pipe, more people, new development, etc.) that might explain the ambiguous outcomes we're seeing. Generally the evidence for an "external factors" explanation appears weak.

In practice, there are some important gaps in program design and implementation that limit the program's effectiveness. Among all the findings from the evaluation, these are probably most relevant in helping to explain the gap between expected and actual safety outcomes:

- *Much of the visible risk is outside HCAs targeted by the program*. The *potential* consequences are clearly higher in HCAs, but 44% of injuries, 46% of property damage, 55% of fatalities, and 69% of the high consequence incidents have been outside the regulatory scope of the program (not in HCAs) during the years IM has been in place. So the potential to impact overall safety outcomes has been somewhat limited. PHMSA has proposed expanding the program to non-HCAs to address this.

- *The risk models widely used to target risks do not reflect the best science on risk analysis*. They use data; they are not driven by data. While the IM risk models follow industry standards and PHMSA's regulations, many and perhaps all are logically and structurally flawed—leading to substantial errors in representing the magnitude and relative ranking of risks. The sample risk model in the liquid IM rule illustrates and reinforces all these flaws. So risk evaluations that would ground an operator's entire risk management program provide faulty information for decision making. The models can't be fixed with minor tweaks.

- *Some of the data going into the risk models introduce errors*. Operators assembled records to implement IM, but for some systems there appears to be a chronic and perhaps growing problem of missing records for older pipe as assets change hands. Data from assessments do not account for all the uncertainty associated with the inspection tools. The requirements for evaluating data quality are vague, imperfect, and cannot be tied to good statistical practices.

- *The repair criteria for defects leave almost no safety margin*. The criteria are based on industry standards that have not been supported or validated by independent analysis. As a result, the rules permit high pressure pipelines in highly populated areas, with known defects and significant uncertainties in the assessment tools, to operate up to 91% (gas) or 98% (liquid pipelines) of the pressure at which the pipe is predicted to burst or fail without requiring an immediate repair. The original design standard provided a safety factor of at least 1.4x to 2.5x in these areas; that safety factor may be effectively erased with the IM repair criteria.[8]

  PHMSA does not capture failure data in a way that would show whether this has contributed to a failure or not. The real risk can't be determined; and that is a serious shortcoming by itself.

- *Much of the machinery for managing risks has not materialized*. Some assessment methods proved difficult for operators to execute correctly. Risk evaluations often were not updated as circumstances changed. Data were not integrated effectively, as accidents sometimes revealed known problems that were not addressed. Preventive and mitigative actions didn't happen to the extent expected. Performance measurement and program evaluation were more limited than expected, and there is no external accountability for performance goals. Pipelines were assessed and repaired, but the broader assessment and management of risks has not really occurred as expected. PHMSA has acknowledged these weaknesses and has proposed several measures in new rules to try to address them.

- *Operators' IM programs are difficult to inspect.* The risk management processes themselves require a high level of specialization and extensive data. Inspectors do not have the expertise in risk modeling, data quality, performance measurement, management of change, and other central elements of a safety management system, to evaluate the reliability of these elements of an operator's IM program. And the measures to judge performance often are not clear. Inspection protocols make extensive use of qualifiers (like sufficient, reasonable, and technically-justifiable) about processes that inspectors often don't have the time, experience, or specialized expertise to judge.

- *The performance-based requirements in IM are difficult to enforce,* and enforcement effectiveness may have been inhibited by long processing times—often many months to initiate a case and years to resolve it. In the meantime, some violations can remain uncorrected. Many inspectors see the enforcement process as unnecessarily burdensome. As a result, the enforcement process is often not the tool of choice for dealing with deficiencies. This probably undermines the deterrent effect of enforcement.

- *Program data/indicators are not providing useful feedback to help guide or redirect the program*. As an enterprise, the program office doesn't know enough about what causes systems to fail, where these conditions exist in the system, how these conditions are changing over time, what barriers are effective in reducing risk, or the status/condition of those barriers. Data collection is not guided by a comprehensive evaluation of *what we need to know* to manage the program effectively—the starting point for any good system of indicators, as outlined in DOT's Information Quality Guidelines.

Several other design and implementation issues might affect safety outcomes.

- While the rules were intended as a supplement to the pipeline safety code, some—maybe many—operators did not add resources to meet the additional requirements.

- Repairs and maintenance, in fact, might be adding risk into the system. Any time a system is disturbed for inspection, maintenance, or repair there is some increased risk. More data are needed to monitor and understand this effect.

- Safety culture—while beyond the scope of IM regulation—is generally understood to be important to achieving safety outcomes but there is no mechanism to measure or monitor it, and there are many recognized counterincentives that could undermine it.

There is a scientific basis for measuring safety culture, with survey scales supported by research; this could be explored further.

- The transition to risk-based, integrated inspections is complex and controversial. There has been no evaluation of the effectiveness of the current (regular) approach before moving to a new one, nor is there an evaluation design built into the new program. The risk model it relies on does not currently provide a sound basis for risk comparisons. And most inspectors I interviewed don't support it, which could undermine effective implementation. Decoupling the *integration* of multiple inspection types from risk-based *targeting* might offer a useful path for addressing these shortcomings.

Several things are evidently working well, and these can provide a foundation for program improvement.

The technology of in-line inspection (ILI) tools has advanced to detect defects that would not have been detected before, and these advances in technology were probably driven by the IM program requirements. Operators are inspecting their systems to a greater extent, and in some areas inspectors are noticing fewer big failures or accidents. Operators are finding and fixing thousands of defects that might have failed. Reassessment intervals appear to be working effectively with repair criteria, with fewer defects appearing over time.

Companies are managing older pipe to safety standards equivalent to newer pipe. Companies are assembling records and learning more about their systems. IM has created a common expectation that operators' safety programs should target the highest risks. And some companies have ventured more deeply into risk management; these might provide useful lessons learned for other companies.

Some observations on the findings and conclusions

Over the years, PHMSA and the pipeline industry have identified many of the immediate "engineering" causes of pipeline failures and worked them out of the system. The long term trends in safety outcomes reflect these advances. But IM was aimed at a deeper level of risk.

It's increasingly well-recognized that major accidents arise from the unforeseen interactions of human and organizational factors—leading to what is sometimes called an "*organizational accident*." The pipeline safety program, like most other federal safety programs, doesn't have data on these kinds of failures because it doesn't have a good conceptual model for them. It investigates relatively few of them. The risk models don't address them. And performance measures don't capture what we need to know about IM assessments, repairs, and risk factors.

Integrity Management is one among a broad class of government safety programs commonly referred to as Performance-Based Regulation (PBR). PBR is partly a response to the risks of organizational accidents. It focuses heavily on things like leadership, processes, continuous improvement and management of change, program evaluation, management reviews, authorities and accountability, and organizational culture. Putting these elements in place requires understanding the drivers (psychological, economic, sociological, physical processes) by which change comes about, and designing strategies and incentives to help ensure that the program is implemented well.

These are not engineering specialties. Designing, implementing and evaluating these programs requires expertise in organizational behavior—which the agency does not have. And there is little information on operators' expertise in this area.

The agency relied heavily on industry for data, standards, technical expertise, and support in designing and implementing IM. These are partly practical constraints related to resources, statutory authority, and administrative procedures agencies must follow. They are also partly a reflection of the industry's own movement into IM standards at the time.

We don't really know what would have happened without integrity management. Perhaps the best way to answer this question for any program is to build program evaluation into the design and implementation. This can provide a strong analytical foundation, testing of assumptions, well thought-out metrics, and evaluation plans that are built into program design. Next best is to build evaluation into redesign and reengineering, which is where the program is today.

It's commonly acknowledged that PBR requires a long period to mature—often decades—to change systems, processes, culture, and organizational behavior; and to work through a process of adaptation and continuous improvement. This evaluation, in fact, was long-planned as a part of this process. In effect, this is applying one of the principles of IM and quality management systems in general—an audit or evaluation process—to the way PHMSA manages the program. The next step, of course, is to act on the findings.

The core logic of the IM program appears to be sound. An argument could be made that if operators were applying it fully and responsibly, the risks would be reduced. But that argument inevitably leads to a further discussion of design, incentives, culture, guidance, oversight, and monitoring. To get different results, some things have to change.

A framework for improving the program

Several changes are already being considered.[9] But in making broad changes in the IM program, expanding the use of integrated inspections, and trying to implement safety management systems, there are a few things that seem important in establishing a sound base for improving the program:

1. *Building much more substantial expertise in the areas needed to oversee performance-based regulations*—to broadly re-tool the IM rules and the inspection and enforcement processes in a way that more clearly considers how the program influences behavior.

2. *Building a much stronger analytical capability*—including expertise in social science, economics, risk evaluation, statistics, and program evaluation.

3. *Developing a good conceptual model of failures*—to underpin PHMSA's data collection, inspection, investigations, and rulemaking programs.

4. *Building a robust information system*—including overhauling the data collection program and metrics, starting with the question: *What do we need to know?*

5. *Correcting the rules and guidance on risk factors and risk modeling*—to fix known problems that are undermining effective safety decisions by operators.

6. *Re-shaping the inspection program into a more forensic, investigative approach*—to adapt to the special challenges in performance-based regulation.

7. *Expanding the accident investigations program*—to provide the primary feedback loop into the program for learning about failures, assessing program effectiveness, and redirecting effort.

8. *Developing a system for managing change*—grounded in credible analysis, testable assumptions, and broad input; and with evaluation *built into* program design.

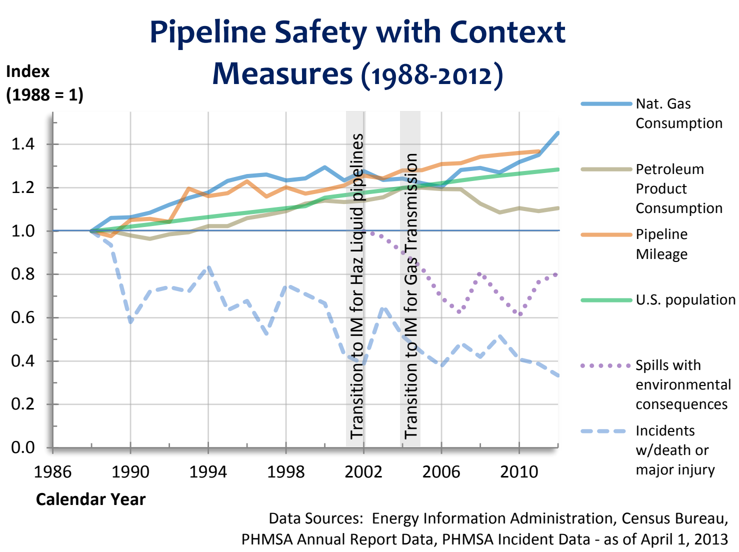
Context – *what shaped the program*

Pipeline infrastructure and safety

Pipelines move most of the primary energy that is consumed in the U.S.—including nearly all of the oil and natural gas that is used to power transportation vehicles and supply the national energy grid, heat homes and workplaces, and provide feedstock for industrial processes and materials.

Transmission pipelines move these gas and liquid products across the country, either directly to industrial users, to temporary storage, or to local distribution systems for further transport. The pipelines are constructed of steel, ranging from 2-48 inches in diameter, and they are mostly buried underground, using an array of methods for protecting them from corrosion and external damage. About 450 operators manage 180,000 miles of hazardous liquid pipelines, and about 1,000 operators manage 305,000 miles of natural gas transmission pipelines, for a total of nearly half a million miles.[10]

Pipelines arguably are the safest, and for natural gas the only practical, mode of transportation for these kinds of hazardous materials, particularly given the enormous quantities and distances they are moved. There are of course inherent risks with hazardous materials. But the number of accidents or incidents[11] with death or major injury has declined an average of 10% every 3 years over the past 27 years, while most measures of risk exposure have increased. Fatalities and injuries from pipeline accidents have declined to 12 and 56, respectively, in 2012, the most recent full year of data. And most (about 4/5) of these serious incidents occur in local gas distribution systems, not the liquid or gas transmission pipelines that have been subject to the IM rules over the past decade.



States provide inspection oversight for about 36% of the pipeline mileage (13% of the hazardous liquid mileage and 49% of the gas transmission mileage) subject to the IM rules today.[12] PHMSA's federal inspectors provide inspection and enforcement oversight for the rest. PHMSA regulates pipeline systems and operators with a budget of about $110 million (FY

2013) including $36 million in grants to States to operate their pipeline safety programs. About 135 federal inspectors and another 300 State inspectors across five regions carry out the field inspection and enforcement program.

PHMSA developed the IM regulations and applied these successively over the past ten years to each of the major sectors of the pipeline industry—hazardous liquid (in 2001), gas transmission (in 2004), and gas distribution (in 2011) pipeline systems. The first two of these, where the program has at least 8 years of operating experience, are the subject of this evaluation.

The history

The IM program built on several earlier efforts, including the Risk Management Demonstration Program, the Systems Integrity Inspection Pilot Program, and a high impact inspection format that evaluated systems as a whole. All of these efforts might be traced to a more general, growing interest in quality management at the time, and a recognition that more integrated approaches to safety and environmental protection would have value.

After some major pipeline incidents in the mid-1990s,[13] regulators began to formulate a new, risk-based approach to pipeline safety, drawing on the experience of the nuclear power industry and several large operators of pipeline systems. These efforts all appeared to offer a way to better focus resources in areas where a pipeline failure could have more significant consequences.

While different companies were using different approaches, the agency began assembling ideas into a single approach that might work for all pipeline systems—a concept that became known as *risk management*. Under this approach, operators would identify and address the most significant integrity threats to their systems, and PHMSA would focus its inspections on operators' integrity management processes and activities instead of a "checklist" approach.

As risk management was being developed, a liquid pipeline rupture and fire at Bellingham, WA (1999)[14] killed three people and led to accelerated efforts to publish an integrity management (IM) rule. The first effort focused on hazardous liquid pipelines. The American Petroleum Institute (API) began developing an IM standard (API-1160), and the agency participated in development of this standard. But the rulemaking was on a faster track. The final rule for hazardous liquid pipelines, published in December 2000, used many of the repair criteria from earlier, draft versions of API-1160.[15]

In June 2000, a natural gas pipeline explosion in Carlsbad, NM[16] killed 12 people and prompted the agency to accelerate its work on an IM rule for gas transmission pipelines. The Bellingham and Carlsbad incidents together led to legislative changes in 2002 requiring integrity management.

Beginning in 1996, the statutory authorities and mandates established a general framework for integrity management, reflecting program initiatives that were already underway at the time.

The Accountable Pipeline Safety and Partnership Act of 1996 (P.L. 104-304) introduced risk management plans, and directed the Secretary to establish risk management demonstration projects to demonstrate the applicability of risk management for liquid and gas pipelines. PHMSA established these demonstration projects, and published the first IM rule—for most hazardous liquid pipelines—under this authorization.

The Pipeline Safety Improvement Act of 2002 (P.L. 107-355) required operators of regulated natural gas pipelines in high consequence areas to conduct risk analysis and implement integrity management programs similar to those already required (by regulation) for oil pipelines. The law authorized the Secretary to issue regulations for integrity management, but the statute itself required natural gas pipeline operators to develop and implement integrity management plans.

Program objectives

One of the most basic elements of a program evaluation is to examine how well a program is meeting its objectives. Sometimes these objectives take shape only through program development and implementation; they are not always explicit when a program is conceived. Often, they are simply "commonly understood." Program evaluation usually begins with an effort to reconstruct program objectives in a form that can be used as criteria for evaluation.

The statutory mandates for integrity management provided only very general guidance on the purpose or objectives of the program. In the 1996 Act, the stated purpose of risk management was *"to protect employees, the general public, the environment, and pipeline facilities."* In the 2002 Act, the stated aim of risk analysis and integrity management programs was *"to reduce the risks."* Notably, OPS published the first regulations for integrity management—for most hazardous liquid pipeline operators—in December 2000, well before Congress mandated integrity management programs in the 2002 Act.

The target problem: The Final IM Rules for both hazardous liquid and gas transmission pipelines identify the target problem in very similar ways (minor differences highlighted in brackets), with a common aim:

> Hazardous liquid pipeline spills [Natural and other gas pipeline breaks can result in explosions and fires that] can have an adverse impact on human health [and safety] and the environment. The magnitude of this impact differs. There are some areas in which the impact of a spill [pipe break] will be more significant than it would be in others due to

> concentrations of people [near the pipeline] who could be affected or to the presence of environmental resources that are unusually sensitive to damage. Because of the potential for dire consequences of pipelines in certain areas, these areas merit a higher level of protection. The OPS is promulgating this regulation *to afford necessary additional protection to these "high consequence areas."*

The rules, as first published for liquid pipelines, were intended "to reduce the potential for hazardous liquid pipeline failures that could affect populated and unusually sensitive environmental areas and commercially navigable waterways."

The basic principles underlying the IM program are that pipeline operators should have a good understanding of their own systems, particularly the threats and risks, and should manage those risks in a systematic way.[17]

Explicit objectives: As the Office of Pipeline Safety (OPS) developed and implemented the IM rules, it began to outline four major objectives in program descriptions and guidance that has been widely published, including on the PHMSA website:

- Accelerate assessments of pipelines in areas with the highest potential for adverse consequences (high consequence areas, or HCAs).
- Promote rigorous, integrated and systematic management of pipeline integrity and improve operator integrity management systems.
- Enhance governmental oversight of company integrity plans and programs.
- Increase public confidence in the safe operation of the nation's pipeline network.

The more "commonly-understood" objectives of the program are to:

- Reduce the number of accidents/incidents (or pipeline failures) in HCAs
- Reduce the consequences (harm to people, environment, and property) of incidents in HCAs
- Reduce the overall consequences of incidents as a result

These objectives are measurable and they are outcome-oriented. They are logically tied to the statutory mandates and program descriptions, and they have obvious public value. They would be accomplished by systematic management of risk, including especially assessments of pipelines that could affect HCAs, with strong governmental oversight. And if successful, one could expect increased public confidence in pipeline safety.

The first two explicit objectives were measured during program implementation; the other two have not been measured. None of the commonly-understood objectives have been measured by the program before this evaluation.

Starting conditions and assumptions

Every program is developed in a context—a set of conditions and general assumptions that shape the objectives, design, and implementation of the program. These can change over time, but the starting conditions and assumptions are important because they establish the baseline for evaluating the original design and for evaluating the effect of changes over time. Generally, these will drive or underpin the general logic of the program, and set the stage for change.

IM development followed from several facts, beliefs, arguments, or indications at the outset:

- *Pipeline incidents with death or major injury had been declining* – since 1986, we have seen a steady decline of these more serious incidents at a rate of about 10% every three years, but the actual numbers of these incidents reflect diminishing returns. This trend suggested that continuing existing programs would not be enough to achieve substantially reduced risk in terms of harm to people.

- *Risk exposure was expected to rise over the next 20 years*, as population increases, demand for energy increases, and new construction encroaches on existing pipelines.

- *Systems are complex* – Pipelines are part of more complex systems, with imperfect design and construction, operated by humans who sometimes make mistakes, and subject to external conditions and forces that can stress the systems.

- *Risk information is incomplete* – Most pipelines are buried (where they cannot be inspected visually), internal and external inspections might be limited for many systems, pipeline operators have/had varying levels of understanding of their systems, and data from many different sources is/was often not integrated.

- *Risk is concentrated geographically* – When pipeline incidents have occurred, most harm to people, property, and the environment was in high consequence areas (HCAs)—areas with significant population, water, or other sensitive natural resources.

- *Many failure conditions are detectable in advance* – about 40% of incidents have been associated with corrosion or material/weld failure, and some additional fraction have been associated with prior damage – the kinds of defects that are often progressive over time, and could be detected with assessment methods and technologies that are available.

- *Some pipeline operators were already managing their systems with IM-like programs*, and the agency's prior work on risk management highlighted the general principles and logic that might be applied to all pipeline systems.

Program logic – *how the program works*

The general logic or theory behind the program

The logic of Integrity management can be distilled into five general elements,[18] each resting on some assumptions:

- *Build risk management processes* – If operators develop their risk evaluation and risk management systems, they will better understand their systems, and be able to prioritize and manage risk in a systematic way.

- *Target HCAs* – If operators focus prevention efforts in the geographic areas where risk exposure and potential consequences of a failure are greatest, the harm to people and the environment will decline at a faster rate. This element aims to take advantage of one key circumstance: that risk is concentrated geographically.

- *Find and fix the most serious defects* – If operators assess their pipelines, they will detect defects or anomalies that might otherwise remain undetected. If they fix the most serious defects first, and reduce operating pressures where an imminent threat cannot be mitigated right away, they will avoid more failures. This element follows directly from the state of technology: that many failure conditions are detectable in advance.

- *Repeat the assessments* – If operators repeat their assessments at least every 5-7 years, they will catch new or progressive problems before the pipeline fails from those causes. Reassessments are needed because systems are complex and some conditions can deteriorate over time.

- *Integrate risk data, evaluate all risks, and mitigate the most serious risks* – If operators integrate all the data they have, they would have a better understanding of their systems. If they evaluate risks with all this information, their risk assessments would better target their resources to the most serious risks. This element is intended to help address the circumstance that risk information is incomplete.

Program interventions

Build risk management processes

Target HCAs

Find and fix the most serious defects

Repeat the assessment before the pipeline fails

Integrate risk data, evaluate and mitigate the most serious risks

Program assumptions

Given the overall objectives and program logic, the success of the program rests on several additional assumptions:[19]

- *HCA criteria* target areas that are actually highest risk re: harm to people and the environment. These criteria are set by regulation, and PHMSA has provided pipeline operators with links to the geographic data that operators use.

- *Enough of the overall risk is concentrated in HCAs* to affect the overall risk and to justify substantial investment of resources in the IM program.

- *Assessment technology* is sufficient—or would be improved sufficiently as a result of the rules—to detect and characterize defects or anomalies accurately.

- *A 5- to 7-year reassessment interval* is sufficient to detect progressive deterioration before a pipeline fails, and is a reasonable length of time to re-check other hazards or threats to the system that might develop.

- *Repair criteria* are complete and targeted to the actual highest risks, and will arrest progressive deterioration before failure or the next assessment. These criteria are established by regulation, drawn in part from a set of industry consensus standards published in API Recommended Practices and the ASME Code.

- *Operators' skills and resources* are adequate to identify systems that could affect HCAs, conduct assessments, interpret and integrate data, evaluate risks, and mitigate the most serious risks.

- *Operators have data from many sources that could be integrated,* but (until IM) they had not integrated it fully. Integration would provide useful new insights on risk, and mitigating these risks would further improve safety.

- *A system-wide approach* is a more effective, and in most cases, more efficient means of inspecting pipeline systems and evaluating pipeline integrity compared to inspecting small segments of systems.

- *One size does not fit all …*the increased flexibility that comes with a performance-based regulation should permit adaptation to fit the unique conditions in each pipeline system and encourage development and use of new technologies.

- *PHMSA's inspections* for the IM program would require a change in approach—from inspecting for compliance with specifications to auditing processes. The agency understood that this requires a somewhat different set of skills, and developed an extensive set of inspection protocols and inspector training to help make this transition.

Other programs to help achieve pipeline safety

The IM program was developed and implemented in an existing and very dynamic program environment. It is working in concert with many other ongoing and new initiatives.[20]

- *Base program requirements* for pipeline safety were largely unchanged by the IM program. Integrity management added new requirements. Of course, this does not mean that operators and regulators simply added new resources to do this work. Many people working in the IM program were drawn from existing programs. To some extent, IM diverted resources from other safety programs, as it was assumed that this was a higher priority. But other programs, including all other kinds of inspections, continued.

- *State programs* – States inspect and enforce compliance with safety regulations for nearly all intrastate gas pipelines and (by agreement) they inspect some interstate gas and liquid pipelines. This includes oversight of about 49% of the total gas transmission pipeline mileage, and about 13% of the hazardous liquid pipeline mileage. The Federal government provides grant funding to states to help defray some of the program costs.

- *Research & Development* – R&D funding is used to improve pipeline inspection technology and analysis tools and strengthen industry's ability to effectively manage pipeline integrity. Research aims to improve operators' ability to prevent damage to pipelines, detect leaks, improve in-line inspection technology, improve oversight of operations and control functions, and access and select stronger pipe materials. PHMSA also studies promising technologies and processes to more fully incorporate risk-based approaches into pipeline operations.

- *Damage Prevention* – As the U.S. population grows and shifts geographically, there is an increased exposure to risk for people living and working closer to pipelines. Commercial and residential development, in particular, increase the probability of excavation damage to pipelines. This has been a top cause of pipeline incidents with death or injury. PHMSA provides grants to States to develop effective damage prevention programs meeting certain minimum standards, with an emphasis on expanding the use of civil enforcement authority against parties who violate "one-call" laws. PHMSA has supported the Common Ground Alliance and played a major role in implementing "811: Call before you dig"—a simplified national program for locating and marking all underground utilities to reduce excavation damage.

- *Targeting operator performance* – Each of the major sectors of pipeline operations presents some unique characteristics. But at least one thing they have in common: a concentration of serious incidents with a relatively small number of operators. PHMSA has focused special attention on operators who have demonstrated weak performance. This includes executive performance reviews to address systemic issues to the highest levels of the operating companies, and plans to reformulate inspection priorities under an "Inspection Integration" initiative.

- *Community Assistance and Technical Services (CATS)* – CATS safety engineers in each region help facilitate communication among all pipeline stakeholders, including the public. They help states assess their damage prevention programs, broaden public awareness of pipeline systems, and facilitate permitting processes.

- *Permitting Assistance* (removing impediments to repair pipelines) – As the pipeline industry implements IM program requirements, the use of new detection technologies is revealing more pipe defects in high consequence areas than previously observed. Hazardous liquid pipeline operators reported at least twice as many defects per year needed timely repair as they did before IM. PHMSA and other federal, State and local permitting agencies conduct many consultations to facilitate repairs for pipeline segments requiring thousands of federal, State and local permits.

- *Transmission Pipelines and Land Use Planning*: PHMSA has worked with property developers, local governments, and the pipeline industry to develop best practices and establish safe land use standards for pipeline maintenance, construction and development in proximity to populated areas.

- *Information grants to communities* – PHMSA provides grants to promote local community awareness and understanding of pipeline safety activities, to promote public participation in pipeline safety proceedings, and to help prevent unintended damage to pipelines.

- *Oil Pollution Act Implementation* – PHMSA protects people and the environment by ensuring that pipeline operators can respond to significant oil spills. This funding goes to: 1) review and approve operator spill response plans; 2) oversee field and table-top exercises to strengthen operator readiness to respond to oil spills from pipelines; 3) monitor major spills and clean-up efforts; and 4) maintain access to information on the location of unusually sensitive areas.

- *Operator Qualification (OQ)* – The PSIA of 2002 required PHMSA to develop qualification standards for pipeline operator safety personnel, inspect operators against those standards, evaluate the effectiveness of the standards, and report the results of those regulations. Additionally, National Transportation Safety Board recommended that PHMSA take steps to ensure that pipeline workers have been trained and are competent to perform their tasks. OQ regulations require operators of gas and hazardous liquid pipelines to conduct programs to qualify individuals who perform certain safety related tasks on pipelines.

- *Control Room Management* – Regulations address fatigue, the man-machine interface, data accuracy for controllers, and qualifications and training for people working in pipeline control rooms.

- *Information collection and analysis* – PHMSA collects data on pipeline systems and accidents or incidents affecting those systems, and has proposed a National Pipeline Information Exchange to develop a comprehensive database of pipeline safety information that integrate information from PHMSA, States, industry, and other Federal sources.

- *The Secretary's Call to Action* – After the pipeline explosions in San Bruno, CA and Allentown, PA the Secretary issued a "*call to action*" and an action plan to accelerate rehabilitation, repair, and replacement programs for high-risk pipeline infrastructure and to re-qualify the integrity of that infrastructure. Followup actions include risk assessments, research, rulemaking, advisories, training, sharing best practices, and public awareness efforts.

External factors that could affect outcomes

Many factors affect the program's ability to achieve its goals and safety outcomes, including many external factors outside the program's control.[21] These include:

- *New construction, changes in HCAs, and increasing mileage* (particularly in HCAs) would tend to increase risk exposure generally, apart from any changes resulting from the IM program. This was expected when the rules were published.

- *New development around existing pipelines* ("encroachment") would tend to increase the exposed population given the same pipeline mileage. This was expected.

- *Continued aging* of much of the infrastructure would tend to increase risk exposure from time-dependent threats, requiring increased attention to manage these threats. This was not explicitly called out in the rules, but would have been expected.

- *Possible evolution of the technology* for pipeline assessment could increase the ability to detect defects or anomalies. This was expected when the rules were published.

- *Changes in the economy*, the profitability of pipeline operations, the amount of product transported, and associated operating pressures in the pipelines could either increase or decrease risk in indirect ways. This was understood to be a range of possibilities, but there was no way to project anything beyond the general projections of GDP and energy consumption at the time the rules were published.
- *Changes in the industry*, including mergers, acquisitions, joint ventures, and other structural or organizational changes could have either positive or negative effects by increasing the transfer of practices across systems, bringing economies of scale, creating instability in cultures or management systems during transitions, or possibly losing information about pipeline systems when assets are transferred. There was no way to anticipate these kinds of changes.
- *Major weather events*, including the severe hurricanes in 2005, generally introduce low probability high consequence risks that are random and largely independent from the management of pipeline systems. There was no way to anticipate these events (in particular); pipeline operators are expected to account for these risks generally as they design and operate their systems.

Integrity management – a program logic model

Starting conditions

Program outcomes reflect diminishing returns

Risk is concentrated geographically

Many failure conditions are detectable in advance

Systems are complex

Risk information is incomplete

Program interventions

Build risk management processes

Target HCAs where risk exposure is greatest

Find and fix the most serious defects

Repeat assessments to catch problems before the pipelines fail

Integrate risk data, evaluate and mitigate the most serious risks

Expected intermediate outcomes

Number and severity of defects will decline over time, especially in HCAs

Operators will better understand their systems

Risks from other causes will decline, especially in HCAs

Management practices will become more preventive

Final outcomes

Pipelines in HCAs will get safer, with fewer accidents and reduced consequences

Pipelines not in HCAs will stay the same or get a little safer as an indirect effect of IMP

Overall safety will improve; accidents will decline; harm to people, environment, and others' property will decline

General assumptions

Large operators were already doing this.

HCA criteria will target areas that are actually highest risk.

Assessment tools/technology are sufficient to detect, measure, & characterize anomalies.

5/7/15 year reassessment interval is sufficient to detect progressive deterioration before failure.

Repair criteria are complete and targeted to actual highest risks.

Operators have adequate skill and resources to implement the program.

External and other factors that might affect outcomes

Damage prevention programs continue to reduce risks from excavation damage.

Inspection integration and RRIM focus PHMSA's inspections on highest risks.

Possible evolution of technology for pipeline assessment.

New construction, changes in HCAs, increasing mileage in HCAs.

Infrastructure continues to age.

Changes in the economy and changes in the industry.

Major weather events can stress the systems.

Program outcomes – *the expected results*

One of the first steps in evaluating program effectiveness is to determine whether the expected outcomes—*the desired results of the program*—are occurring. The preambles to the rules and the accompanying regulatory evaluations provide a range of expected benefits, or outcomes. There are also many other outcomes implied in the program logic model and stated in various forums over time as the program has been implemented.

To assess effectiveness, we need to understand the nature and magnitude of these outcomes (expected vs. observed), the strength of the evidence, the extent to which external factors might be affecting these outcomes, and whether there are other, unintended effects (positive or negative) from the program. We also need to assess whether the original expectations were reasonable. Ideally, ultimately, we want to develop and evaluate evidence that the *program* is contributing to the outcomes we observe. First, some simple comparisons—what we expected and what we have observed.

Immediate and intermediate outcomes

There are many things program managers expected to happen as a result of the IM rules. These are often called *immediate or intermediate outcomes*—actions, direct effects, changes in behavior and the direct effects of those changes—which ultimately lead to the broader safety outcomes that people expect from the program. Program managers generally watch these kinds of intermediate outcomes to get an early read on how the program is working, to help adjust and adapt the program over time. When these expectations are stated at the time (as in the preambles to the IM rules), they provide good insight into the theory of the program and the rationale for its design.

Through interviews with program managers and review of the rules and other program documentation, several expectations emerged as clearly important in testing the logic and implementation of the IM program:

- *The number and severity of defects* would decline over time, particularly in HCAs, as operators find and fix problems. This should result in fewer incidents involving corrosion, material/weld failure, and failures associated with previous damage.

- *Risk from other causes* would decline, particularly in HCAs, as operators integrate their data and manage risk more systematically.

- *The number and consequences of incidents in HCAs* would decline as resources are concentrated in these areas and as repair conditions are standardized.
- *The number of incidents outside HCAs* might decline incidentally as assessments are done and risk information is better integrated; but incidents might increase to some extent as resources are targeted to areas of higher consequence.
- *Management practices* would become more preventive as operators better understand their systems, as the number of defects declines, and as harm to people and the environment declines.

Expected safety and economic outcomes

For hazardous liquid pipelines, the preamble to the IM rule stated that the agency "does not have adequate data on pipeline spills to accurately gauge the benefits of this rule." The agency concluded that the costs of the rule were justified based on:

- the subjective benefits of improving knowledge of pipe condition, addressing public concerns, and *reducing the frequency and consequence of pipeline releases that affect high consequence areas*.

This suggests measuring the total number of accidents, deaths, major injuries, property damage, and environmental damage in HCAs, or segments that could affect HCAs (this report will refer to these generally as "in HCAs" for simplicity).

For gas transmission pipelines, the preamble to the IM rule and the accompanying regulatory evaluation stated that the direct safety benefits of the rule will be realized in:

- *reducing the consequences of accidents, including deaths, serious injuries, and property damage* (estimated at $800 million in benefits over 20 years)
- *averting accidents with larger consequences than any experienced to date*, because the rule was focused on precisely the high population areas in which they could occur (estimated at $277 million in benefits over 20 years); and
- *averting accidents* more generally.

The regulatory evaluation with the gas transmission rule did not specify "in HCAs" for these consequences, but the logic of the program clearly aims to achieve these reduced consequences overall by focusing in HCAs.

The pipeline accident/incident data provide reasonably good measures of accidents/incidents, deaths, serious injuries, property damage, and (for hazardous liquids) the amount spilled. The data after 2002 also include limited information about environmental damage from hazardous liquid pipeline accidents. But since the rulemakings created the framework for identifying HCAs, there is no historical baseline data (pre-IM) on accidents or incidents in HCAs for comparison.[22] So in this analysis, I examined and compared total numbers before and after IM implementation, adjusting for inflation and changes in reporting criteria, and considering changes in pipeline mileage as an explanatory factor. I also looked at the trends in the data after IM implementation, since we expected to see any improvements emerge over time, not all at once.

The data – *what actually happened*

Improvement in safety outcomes are not occurring as expected

Accidents/ Incidents & Consequences	Hazardous Liquid Pipelines	Gas Transmission Pipelines
Accidents/ Incidents	~	✗
High Conseq. Incidents	~	✗
Deaths	~	~
Injuries	~	~
Property Damage	~	✗
Corrosion Failure	~	~
Material Failure	~	✗

Trends before/after IM implementation

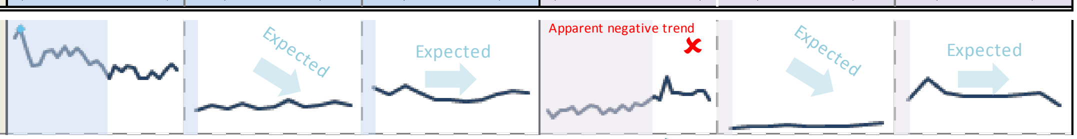
Some things have gotten better,[23] but we have not seen clear improvement in any safety outcomes that the program targeted. Some things (about half of the incident and consequence indicators) appear to have gotten worse since IM implementation, and some of these trends seem to be continuing in the wrong direction. It's difficult to explain these patterns given the program design, focus, and expectations.

To assess the trends, I plotted the incident and consequence data through the last year before IM implementation, developed statistical trendlines to help factor out annual fluctuations and establish a reasonable baseline for comparison, then compared the data after IM implementation to determine if the new trend was significantly better (✓), worse (✗) than the baseline, or simply inconclusive (~). Appendix B provides further details on the statistical methods and limitations of this analysis.

Overall, the patterns in terms of consequences are very different for hazardous liquid vs. gas transmission pipelines.

Accidents/ Incidents & Consequences	Hazardous Liquid Pipelines			Gas Transmission Pipelines		
	Total (1986-2002-2012)	**HCAs** (2002-2012)	**Not in HCAs** (2002-2012)	**Total** (1986-2004-2012)	**HCAs** (2004-2012)	**Not in HCAs** (2004-2012)

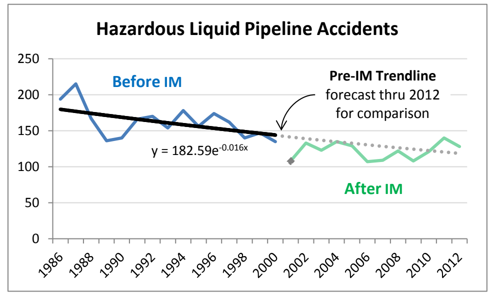
Accidents/ Incidents (High=215)



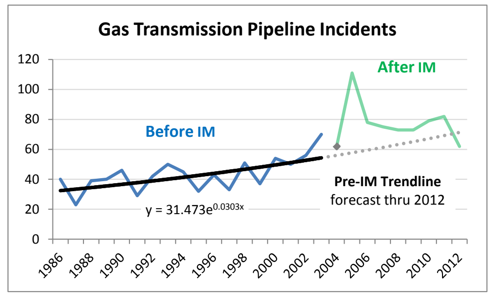
***The total numbers of accidents and incidents** are generally falling for liquid pipelines and rising for gas pipelines, but for both sectors the numbers are essentially flat (or rising slightly) in HCAs since IM implementation.*

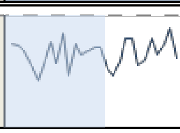
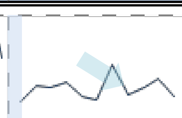
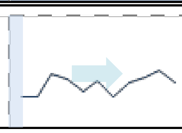
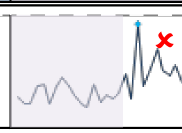
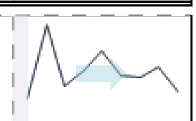
While the IM rules focused especially on the *consequences* of incidents, the regulatory evaluations expected to avert gas transmission incidents generally, and reduce the frequency of hazardous liquid releases that could affect HCAs. The evidence suggests that neither of these outcomes are occurring as expected.

Hazardous liquid pipeline accidents were declining at a rate of about 1.6% per year before IM was implemented. Then there appears to have been *a phase shift*, with fewer accidents in every year after IM implementation but no obvious trend in the new data—i.e., no continuing decline—over the past 10 years. There might be a slight increasing trend in HCAs over this period.[24] The numbers of accidents affecting, or not affecting, HCAs now are roughly proportional to pipeline mileage in these areas, respectively.



Gas transmission pipeline incidents were rising at a rate of about 3% per year before IM implementation and the total number of reported incidents has increased even more after IM implementation. In fact, every year since IM implementation in 2004 has been higher than any of the 18 years before IM. While the numbers and proportion in HCAs are relatively small, the patterns here appear to be increasing since 2004 as well.



Accidents/ Incidents & Consequences	Hazardous Liquid Pipelines			Gas Transmission Pipelines		
	Total (1986-2002-2012)	HCAs (2002-2012)	Not in HCAs (2002-2012)	Total (1986-2004-2012)	HCAs (2004-2012)	Not in HCAs (2004-2012)
High Conseq. Incidents (High=50*)						

High consequence (HC) incidents[25] *are becoming more prevalent, but with different patterns for hazardous liquid and gas transmission pipelines*. One specific benefit expected from the gas transmission IM rule was to avert "accidents with larger consequences than any experienced to date." Clearly this did not occur.

The pipeline explosion and fire at San Bruno, CA in 2010 was undoubtedly the highest consequence gas transmission incident in the 27 years for which we have data. It resulted in 8 deaths (the third highest on record), 51 injuries (the highest on record, equal to the next nine incidents together), and $387 million in property damage (over three times as high as the next highest, adjusted for inflation). This incident occurred in an HCA.

The crude oil spill at Marshall, MI in 2010 was the highest consequence hazardous liquid pipeline incident over the same 27-year period. It resulted in over 840,000 gallons released into U.S. waters, an estimated $835 million in property damage (nearly five times as high as the next highest on record), and un-quantified environmental damage. This accident occurred on a segment that could affect (and did affect) an HCA.

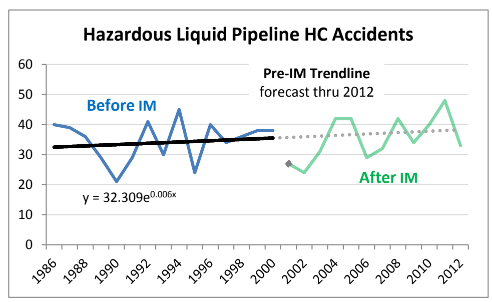
In terms of economic consequences, the top six (and seven of the top ten) liquid and natural gas transmission incidents[26] all occurred since the effective date of the IM rules. Before IM, the incidents at *Bellingham* and *Carlsbad* were the largest consequence incidents on record.

Top 10 liquid/gas transmission pipeline accidents (1986-2012) in terms of economic consequences

Date	Gas or Liquid	City/County	State	Deaths	Injuries	Property damage (2012 $)	Total Economic Consequences	Reported Cause
7/26/2010	Liquid	Marshall	MI	0	0	$834,828,822	$834,828,822	Material failure
9/9/2010	Gas	San Bruno	CA	8	51	$386,868,952	$465,468,952	Material failure
8/30/2005	Liquid	Buras	LA	0	0	$172,293,700	$172,293,700	High winds
7/1/2011	Liquid	Yellowstone	MT	0	0	$137,247,703	$137,247,703	Heavy rains/flood
5/13/2005	Gas	Marshall	TX	0	0	$100,504,658	$100,504,658	External corrosion
2/27/2008	Gas	Hartsville	TN	0	0	$82,913,063	$82,913,063	High winds
6/10/1999	Liquid	Bellingham	WA	3	8	$59,429,198	$82,229,198	Other/miscellaneous
8/30/2005	Gas	Port Sulphur	LA	0	0	$74,666,346	$74,666,346	Heavy rains/floods
8/19/2000	Gas	Carlsbad	NM	12	0	$1,292,294	$73,292,294	Internal corrosion
4/7/2000	Liquid	Prince Georges	MD	0	0	$64,725,000	$64,725,000	Failed pipe

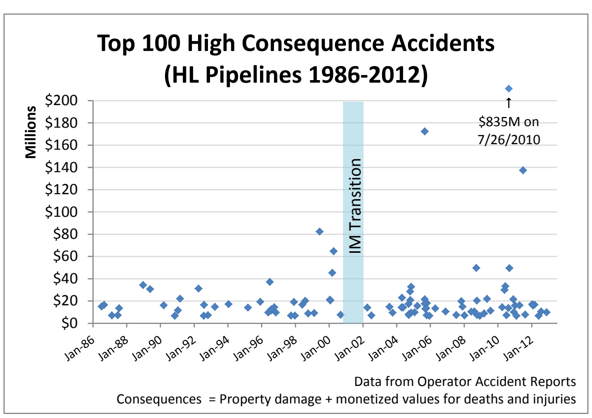
Total economic consequences + Property damage (2012$) + Deaths (@$6M) + Major Injuries (@$0.6M).
Offshore incidents are excluded. Accidents/incidents after IM highlighted in blue font.

More broadly, the trends for high consequence (HC) incidents[27] vary by sector:



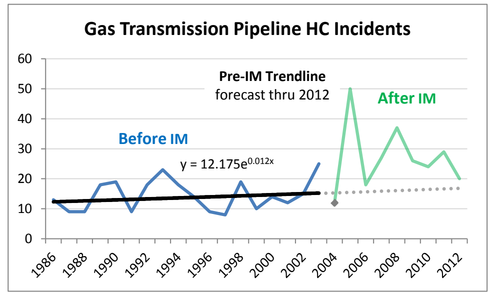
For liquid pipelines, there is no discernible trend in the total number of HC accidents over time. There were 376 HC accidents in the 11 years (1990-2000) before IM implementation and 397 HC accidents in the 11 years (2002-2012) after IM.

The *largest* consequences, however, appear to be increasing—of the top 100 liquid accidents in terms of economic consequences, 41 occurred in the 15 years before IM (2.7 per year), and 59 occurred in the 11 years after (5.4 per year)—as the rate per year has roughly doubled. Interestingly, none occurred during 2001, as operators transitioned to IM implementation.

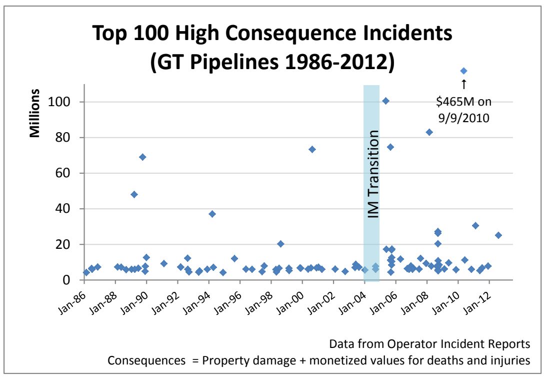


About 13% (944/7,138) of all liquid pipeline accidents from 1986-2012 were HC accidents. These accounted for all 359 casualties, $3.2 billion in property damage (91% of the total, in 2012 dollars), and 76% of all product lost.

For gas pipelines, there has been a noticeable increase in the number of HC incidents since IM was implemented. There were 112 HC incidents in the 8 years (1996-2003) before IM and double that number—231—in the 8 years (2005-2012) after IM. Five of the eight years after IM were the highest on record for HC incidents.



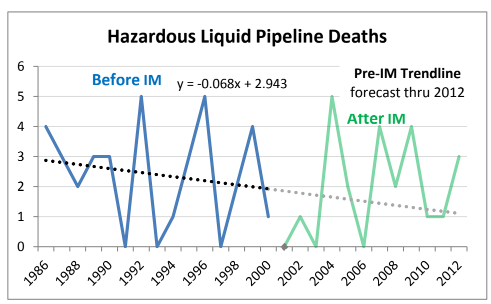
Of the top 100 HC incidents since 1986, 44 occurred in the 18 years before IM (2.4 per year) and 56 occurred in the 8 years after IM (8.0 per year). The annual rate of these very high consequence incidents has tripled.



About 21% (505/2,413) of all gas transmission incidents over the 27 year period 1986-2012 were HC incidents. These accounted for all 388 casualties and $1.6 billion in property damage (87% of the total, in 2012 dollars).

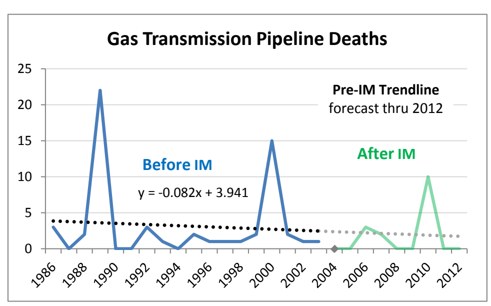
Accidents/ Incidents & Consequences	Hazardous Liquid Pipelines			Gas Transmission Pipelines		
	Total (1986-2002-2012)	**HCAs** (2002-2012)	**Not in HCAs** (2002-2012)	**Total** (1986-2004-2012)	**HCAs** (2004-2012)	**Not in HCAs** (2004-2012)
Deaths (High=22)						

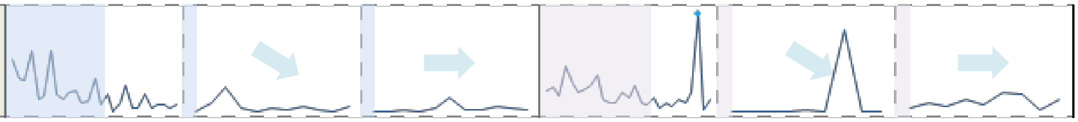
The number of deaths *per year is very low for both hazardous liquid and gas transmission pipelines*—together they account for only about 20% of all regulated pipeline fatalities (most are on gas distribution systems).



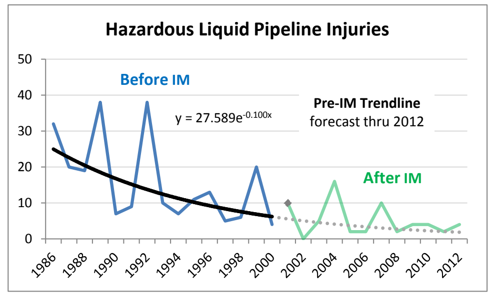
Since IM implementation, there has been an average of about 2 deaths/year for each of these two sectors. This is about the same as the comparable (11-year) period before IM for hazardous liquid pipelines, and about two-thirds the previous average for gas transmission, despite the incident at San Bruno in 2010 that killed 8 people.

The numbers for both liquid and gas transmission pipelines are too small to draw meaningful comparisons before/after IM implementation.

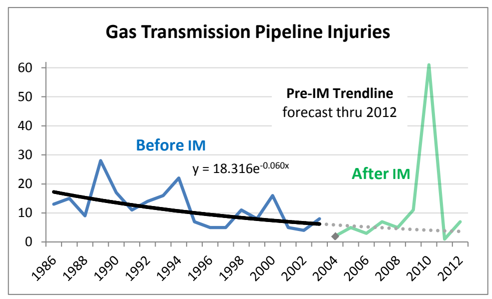


Accidents/ Incidents & Consequences	Hazardous Liquid Pipelines			Gas Transmission Pipelines		
	Total (1986-2002-2012)	HCAs (2002-2012)	Not in HCAs (2002-2012)	Total (1986-2004-2012)	HCAs (2004-2012)	Not in HCAs (2004-2012)
Injuries (High=61•)						

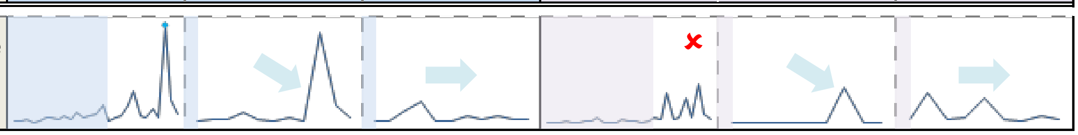
***Major injuries** appear to be declining for hazardous liquid pipelines, but there is no evidence of a reduction for gas transmission, since IM implementation.*



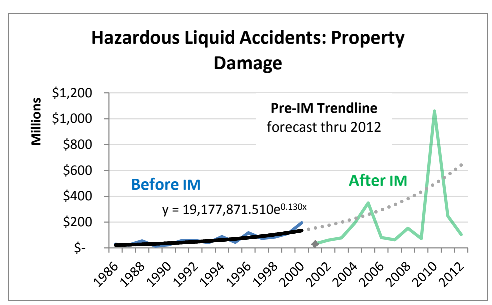
For hazardous liquid pipelines, most years since IM implementation have had fewer injuries than the baseline for comparison (from the 1986-2000 trend). The past eleven years have seen the five lowest numbers (0-2) of serious injuries in the 26-year record; another three years were as low (4) as any previous year.



For gas transmission, four of the eight years since IM implementation had fewer injuries than the baseline, and four had more. Two of the past 8 years have been record low numbers (1-3) of serious injuries. The total numbers of injuries are skewed substantially by the incident at San Bruno in 2010.

Accidents/ Incidents & Consequences	Hazardous Liquid Pipelines			Gas Transmission Pipelines		
	Total (1986-2002-2012)	HCAs (2002-2012)	Not in HCAs (2002-2012)	Total (1986-2004-2012)	HCAs (2004-2012)	Not in HCAs (2004-2012)
Property Damage (2011 $) (High=$1.06M•)						

***Property damage:** Incident reports show property damage overall rising substantially* – clearly in contrast to the expected reduction from IM.

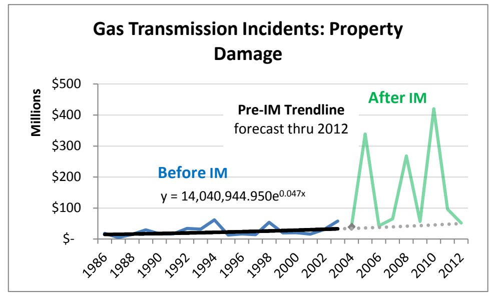


Very large/consequential accidents can drive the overall numbers for any year. But there are also more accidents each year involving large amounts of property damage.

For hazardous liquid pipelines, 31 of the top 50 accidents and 62 of the top 100 (in terms of property damage, adjusted for inflation, over a

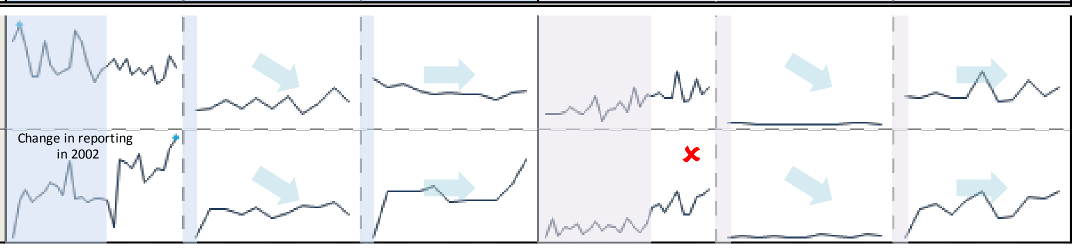
27 year period) occurred in the 11 years since IM implementation. The top 3 ranged from $137 million to $835 million in 2012 dollars, and all of these occurred since 2005; no previous spills resulted in more than $65 million in reported property damages. Since 2002, the average annual property damage was $223 million—*almost three times the annual average* for the previous eleven years.

For natural gas transmission pipelines, 35 of the top 50 incidents (over the same 27-year period, adjusted for inflation) have occurred in the 8 years (2005-2012) since IM implementation. The top four, and 9 of the top 10, all occurred in the last eight years. Since 2004, the average annual property damage was $167 million—nearly *six times* the annual average for the 8-year period before IM implementation.



When a pipeline failure occurs, the potential damage is expected to be disproportionately higher in an HCA, by definition. But since IM implementation, only 5 of the top 100 gas transmission incidents and 49 of the top 100 hazardous liquid accidents occurred in HCAs (or segments that could affect an HCA). These numbers are roughly proportional to pipeline mileage in HCAs (6% and 44%, respectively). This same pattern appears in the overall number of reported (significant) incidents; for both HL and GT the incident rate in HCAs is about the same as the rate not in HCAs.

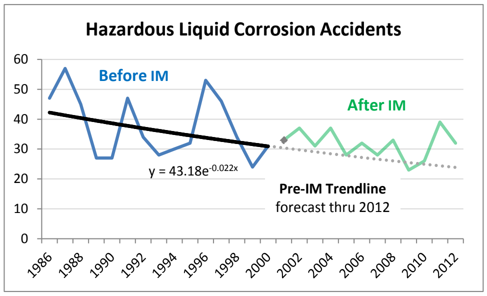
Accidents/ Incidents & Consequences	Hazardous Liquid Pipelines			Gas Transmission Pipelines		
	Total (1986-2002-2012)	HCAs (2002-2012)	Not in HCAs (2002-2012)	Total (1986-2004-2012)	HCAs (2004-2012)	Not in HCAs (2004-2012)
Corrosion Failure (High=57*)						
Material Failure (High=57*)						

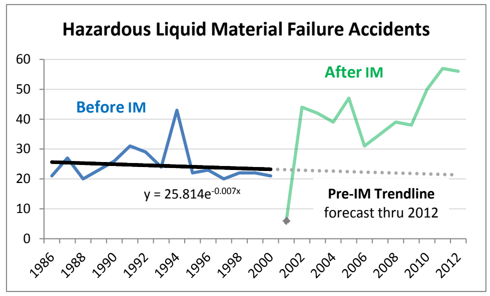


Corrosion and material failure: *The IM program concentrated most attention on two accident causes—corrosion and material failure—in HCAs.*[28] *The program expected these incidents to go down; instead, they are rising for gas transmission pipelines. Data for liquid lines are inconclusive*.

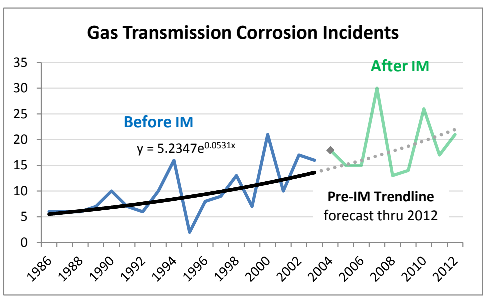
Corrosion and material failure are the two most frequent causes of incidents for both liquid and gas transmission pipelines.

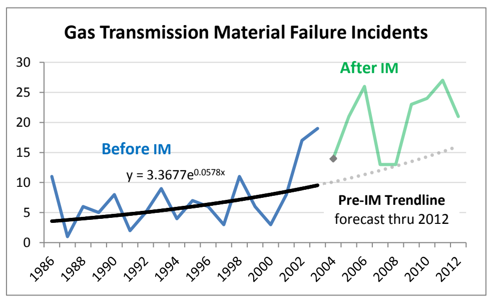
Hazardous liquid pipeline accidents due to corrosion appear to have leveled off after IM implementation—some years higher than the baseline and some years lower. Accidents attributed to material failure clearly increased after 2001, but this coincided with a change in reporting of causes that might account for all of this increase**.** Corrosion and material failure are the most frequent causes of HL accidents, accounting for 23% and 48%, respectively, of all reported HL accidents over the past 11 years. The next highest cause is human error at 11%.

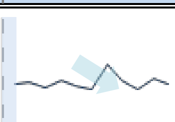
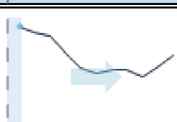
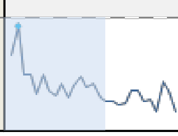
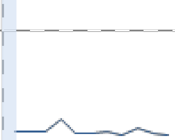
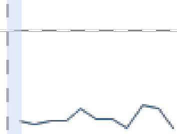




Gas transmission incidents due to corrosion and material failure have increased overall. Both trends were rising before IM implementation, and the numbers have been above the "expected" baseline all but one year since, including the two highest numbers of corrosion incidents on record (30 in 2007, and 26 in 2010), and the six highest numbers of material failure incidents on record (peaking at 27 in 2011). Corrosion and material failure are the most frequent causes of GT incidents, accounting for 20% and 28%, respectively, of all reported GT incidents since 2004. The next highest causes are excavation damage and natural force damage, both at 14%. (The change in reporting might account for some increase here too.)





Accidents/ Incidents & Consequences	Hazardous Liquid Pipelines			Gas Transmission Pipelines		
	Total (1986-2002-2012)	**HCAs** (2002-2012)	**Not in HCAs** (2002-2012)	**Total** (1986-2004-2012)	**HCAs** (2004-2012)	**Not in HCAs** (2004-2012)
Spills w/Environ Impacts (High=108*)	*Data not available before 2002*			*N/A for gas transmission*	*N/A for gas transmission*	*N/A for gas transmission*
Barrels Spilled (High=396,000*)				*N/A for gas transmission*	*N/A for gas transmission*	*N/A for gas transmission*

Environmental harm: Hazardous liquid spills with environmental consequences are generally declining outside HCAs, but essentially flat in HCAs (or in segments that could affect HCAs).

Corrosion (31%) and equipment failure (20%) were the leading causes of these spills. The amount of product spilled overall declined during the 15 years before IM implementation, and the trend since IM is flat or declining.

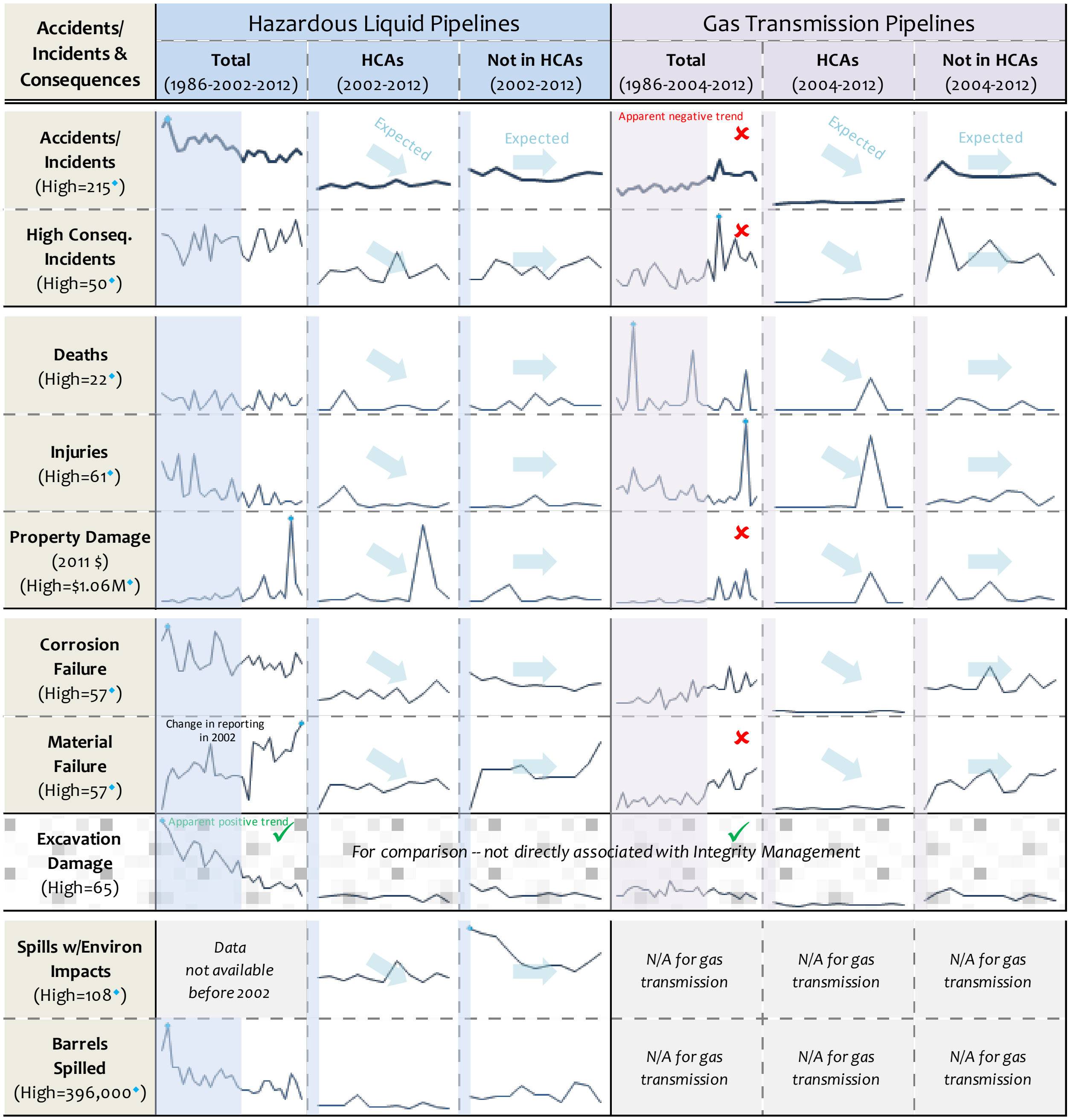
The Integrity Management Progress Reports to Congress in January-February 2011 painted a somewhat different picture of program results, suggesting that "*Some measures developed using recent … accident/incident history appear to indicate that the IM rule is having a positive impact on frequency and consequences.*" There are several reasons for this difference. In evaluating outcomes, this program evaluation:

- adds three more years of data (2010-2012) which show up-ticks on some measures,
- adds some safety outcome measures, like high consequence accidents, that reflect the program logic and explicit expectations from the original IM rules,
- adds an analysis of incidents caused by material failure—the largest single cause of accidents/incidents for both HL and GT pipelines, and one that is particularly relevant for evaluating the IM program,
- uses a method for comparing before/after results that accounts for two separate trends rather than simply comparing averages (many of the differences vanish when accounting for the pre-IM trend),
- uses a type of regression (exponential) that is better (than moving averages) at representing trends for time series data that show diminishing returns, and
- disaggregates each of the measures to determine to what extent the results are in the HCAs that were targeted by the rules.

These progress reports acknowledged the ultimate objective of reducing the likelihood and consequences from releases that could affect HCAs, which is central to the program logic outlined here. They also introduced a new expectation, not articulated in the rulemaking—that results would be observable "over the long term." This is consistent with the experience of many other safety programs that have tried implementing performance-based regulation. But it does not really explain why we are not seeing more substantial results over the 8-11 year experience with IM so far.

Most inspectors, region directors, and program staff were surprised by the safety trends developed in this evaluation. They offered a variety of possible explanations—analyzed and discussed beginning on page 33.

Accident/Incident Trends: *Before and After Integrity Management*



Source: Accident/incident reports submitted by pipeline operators (as of February 22, 2013)

Notes:

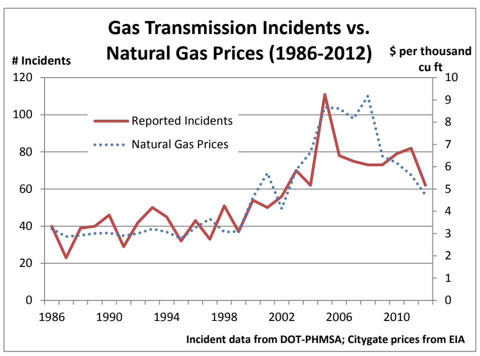
All graphics in each row (e.g., Accidents/Incidents) use the same scale for comparison. These are "significant" incidents.
The range for each row is from zero to the highest ("High") value across both gas and liquid sectors.
HCAs (for Hazardous Liquid pipelines) means any segment that *could affect* an HCA.
Shaded areas indicate years before Integrity Management; 2001 was a transition year for liquids, 2004 for gas transmission.
High Conseq. Incidents include any with death, major injury, ≥ $500,000 property damage (2012 $), or liquid spill ≥ 100,000 gals.
✗ indicates a trend in the "wrong" direction re: expectations; ✓ indicates the "right" direction; all others are inconclusive.

The data above do not account for external factors which might help explain the trends.

Some possible explanations

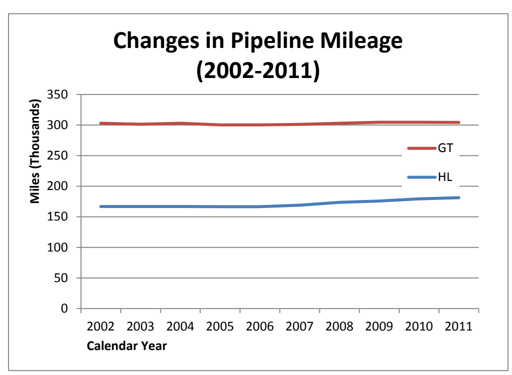
External factors and other unanticipated changes or trends might help explain the outcomes (or lack of outcomes) we are seeing. Some of these are clearly more important than others, and in some cases the evidence challenges conventional wisdom. Collectively, they could help explain some of the apparent increases in adverse outcomes. But they don't provide a very satisfying explanation for why things have not *improved* more convincingly.

Inflation could increase the consequences from accidents/incidents, and might also increase the number of releases that must be reported based on the property damage threshold. But simple inflation is already factored out in the analyses, using inflation-adjusted dollars, and most long-term trends have already adjusted for the effects of inflation on reporting thresholds (using the concept of "significant incidents"). There might be other changes in the value of certain kinds of property, but there isn't enough data to distill the effects of these. General inflation does not account for the outcomes we see.



Natural gas prices rose dramatically from 2000-2005, and generally remained higher than previous levels (1986-2000). Most natural gas incidents are reported because they meet the threshold for property damage, which includes the cost of gas lost. So as gas prices rise, we might expect more reportable incidents as a result. PHMSA does not have disaggregated data on property damage before 2010, so we can't really determine to what extent the increase in reporting is attributable to gas prices. But the patterns (see graph at left) appear to follow each other closely over the past 27 years, supporting this explanation for one trend—the increase in the number of gas transmission incidents.

Pipeline mileage could generally increase risk exposure. But small increases in pipeline mileage[29] are substantially less than most of the trends, and generally less than the increases before IM implementation, so changes in mileage cannot account for more than a very small fraction of the outcomes.

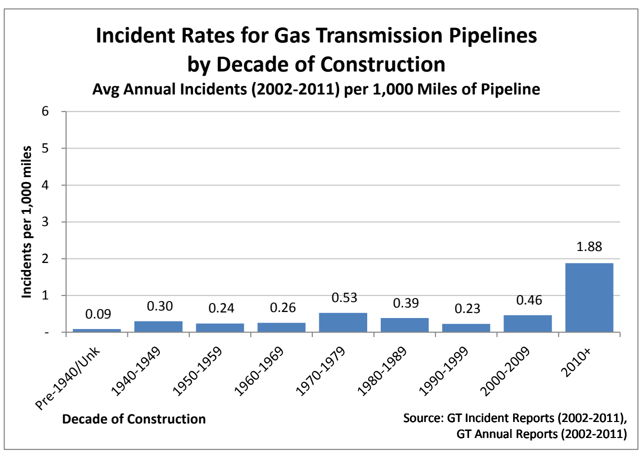


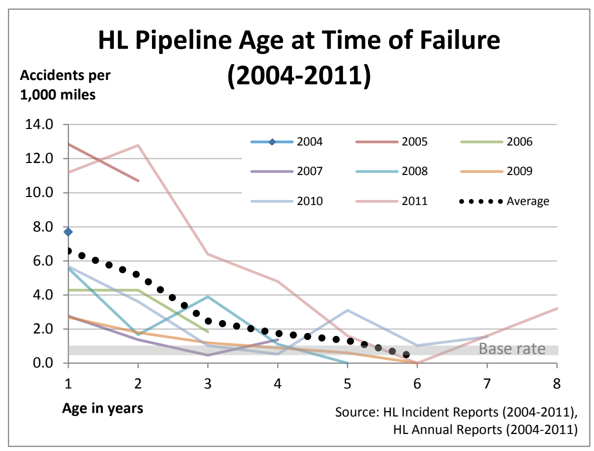
The overall mileage subject to the rules was, in fact, twice as much as estimated[30] when the rules were published, so there should have been greater impact from IM than expected. The agency estimated that 35,500 miles of liquid pipeline would be impacted by the rule; it was actually over 72,000 miles as the rule was implemented. Gas transmission mileage estimates proved much closer to the mark, but only amounted to about 7% of total mileage.

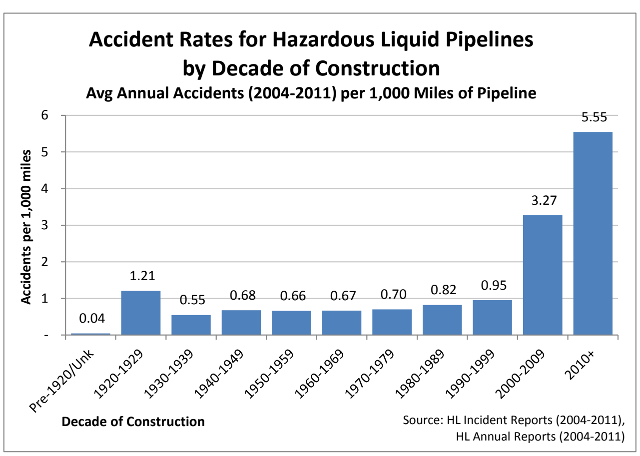
Changes in HCAs: There have been no updates to the environmentally-sensitive areas or drinking water areas since HCA maps were first published in 2001; navigable waterways have been updated every two years with Coast Guard data, but the changes have been very minor; population areas were not modified until September 2012 from 2010 Census data. So changes in the published HCAs have not significantly affected the data since IM implementation.

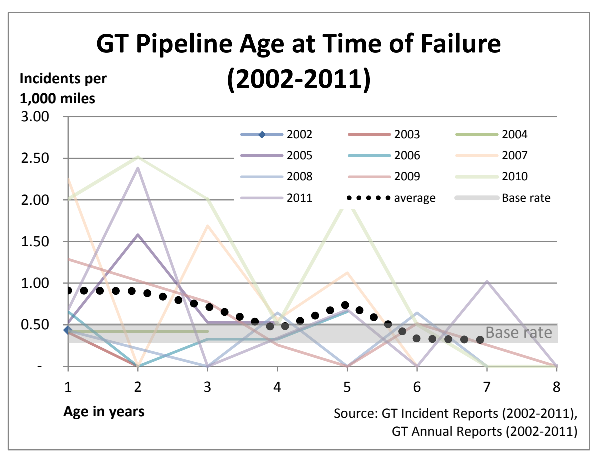
Small changes in risk exposure have probably occurred in fact, as the Census data have shown an increase of about 1.6% per year in the square mileage of populated areas over a 10-year period. And pipeline operators have added new pipelines and in some cases reclassified existing pipelines in areas that could affect HCAs. In any case, the annual reports show a modest 10% increase in HCA mileage for hazardous liquid pipelines from 2004-2011, and a 6% *decline* in HCA mileage for gas transmission pipelines over the same period (PHMSA did not ask for HCA mileage before 2004).

Aging infrastructure: The infrastructure is aging, but the data suggest that pipe 50-80 years old is just as "safe" as pipe that is 10-50 years old. Newer pipe (0-6 years old) tends to present a greater rate of failure, but there is not enough newer pipe to account for the increases in





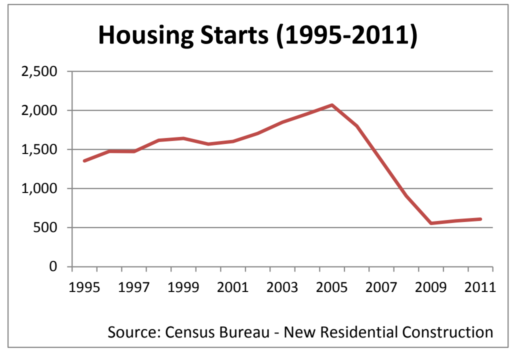




accidents. At about 6 years old, the failure rate for newer pipe reaches the "base rate"—the level we continue to see for pipe up to 80 years old.[31] The only older pipe (by decade of construction) that presents an increased failure rate is hazardous liquid pipe installed between 1920-1929, coinciding with the first construction boom for oil pipelines.

Older pipe might (probably does) still present an inherently higher risk, but evidently it is managed and maintained to a fairly constant level of failure compared to newer pipe.

Increasing development around pipelines can increase the exposed population and present increased risk of damage to pipelines. Development was expanding before IM implementation, but housing starts leveled off then dropped dramatically (over 70%) between 2004-2009[32], and the total urban population in the U.S. grew only 12% from 2000 to 2010.[33] So the effect of new development is likely to be modest, and in fact *more* rapid development was anticipated at the time of IM implementation. In other words, it is likely that this was already built into the expectations.



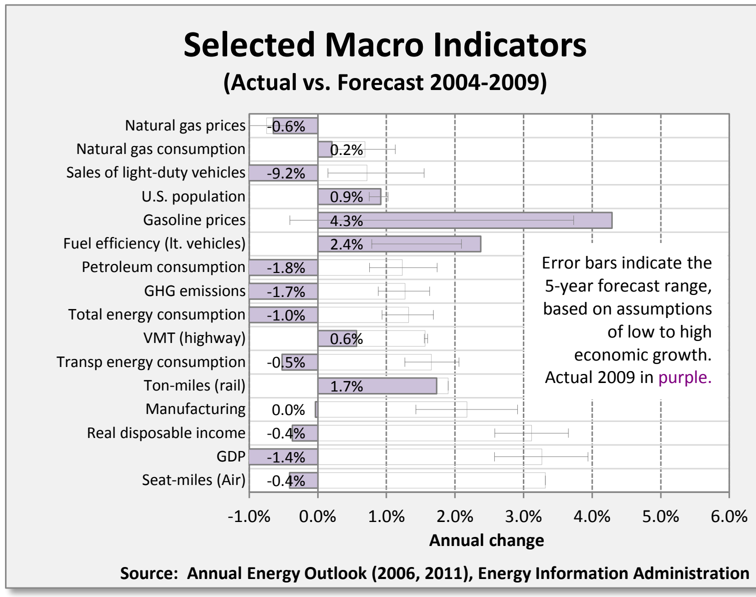
Changes in the types of incidents occurring: Some have suggested that there are changes in the kinds and locations of incidents in recent years, involving more tanks and facilities that are not subject to IM assessment (vs. line pipe); and more spills on operator property (vs. the public right of way). The data are not easy to analyze along these lines, and changes in reporting over time limit our ability to analyze these patterns.

In 2002, PHMSA began collecting data on the system location for each accident, indicating whether the accident occurred in the pipeline right of way (public space), on operator property, in a high consequence area, or some combination of these. But operators were not required to provide the data, and about half the accident reports have no location information.

In 2010, PHMSA revised and simplified the reporting for hazardous liquid pipelines. For the three full years (2010-2012) for which the program has data, about two-thirds (49%) of the significant accidents were totally contained on operator property. These accounted for only about 8% of all property damage reported, but 40% of the deaths and injuries. And in nearly 7/10 of these "totally contained" spills the operator reported impacts to soil, water, fish, birds, or other wildlife.

It is possible that these patterns are changing. In any case, it is not clear why that would make the overall outcomes more negative under the IM program, since IM was intended to be risk-based and performance-oriented, and to address all kinds of risk. Those attributes should make IM *more* responsive to these kinds of changes if they are driving overall risk measures.

Changes in the U.S. economy can affect the production and consumption of energy products as well as the profitability of pipeline transportation.



In the mid-2000s, energy consumption was expected to rise by about 1.3% per year over the next five years. But the economic downturn in 2008 resulted instead in an overall decline in GDP and a 1.8% average annual decline in petroleum consumption (instead of an annual increase of +1.2% projected). Natural gas consumption rose an average of 0.2%/year but this was still smaller than the projected increase of 0.7% per year.

So while a growing economy theoretically might bring increased risk exposure, the actual changes over the past several years cannot explain the outcomes we are seeing.

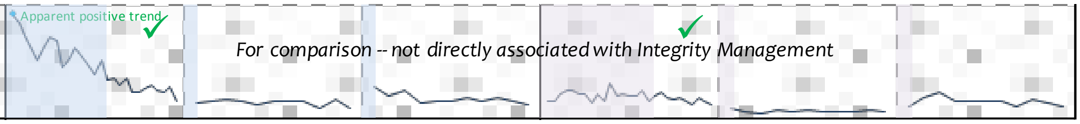
Major weather events, particularly the strong hurricanes of 2005, increased property damage and—to a lesser extent—the number of incidents and high consequence incidents that year. But all of the trends are fairly obvious and consistent even after accounting for the 2005 hurricanes.

Of the top ten liquid/gas transmission pipeline accidents (1986-2012) in terms of economic consequences, four were attributed to high winds or heavy rains/floods. These are not the kinds of risks that would be addressed directly with assessment programs under IM. But the IM program more broadly is not limited in the causes it was intended to address. Operators are supposed to (and do) take actions to minimize the risk from natural force damage. And even more generally, from a public safety perspective, all risks have to be included in an assessment of the potential impacts from the presence of a pipeline.

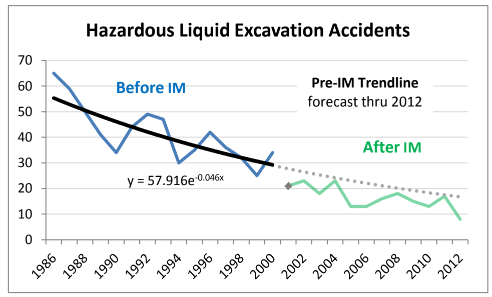
A few very large accidents drive the total amount of property damage in any given year. For example, the top 3 accidents each year over the last ten years account for 27-87% (two-thirds overall) of all property damage on hazardous liquid pipelines in each of these years. This doesn't really explain the trend in property damage, but it does provide some perspective on it.

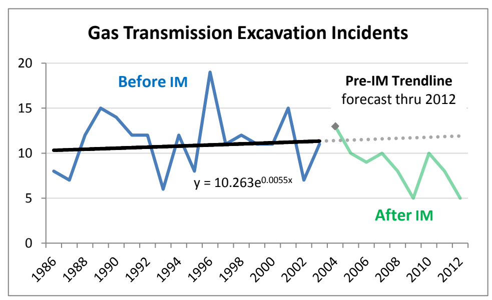
Outliers in general can create some difficulty in analyzing and interpreting the data. At the same time, we need to be very careful about excluding outliers from the analysis, particularly for a program that is focused on consequences. In statistical analyses generally, outliers are sometimes excluded because they indicate measurement error (faulty data), or they cannot be handled easily with the statistical tools being used. But the pipeline safety program, and the IM program in particular, is focused on low-probability high-consequence risks. In fact, they often present a disproportionate level of public concern.[34]

Accidents/ Incidents & Consequences	Hazardous Liquid Pipelines			Gas Transmission Pipelines		
	Total (1986-2002-2012)	**HCAs** (2002-2012)	**Not in HCAs** (2002-2012)	**Total** (1986-2004-2012)	**HCAs** (2004-2012)	**Not in HCAs** (2004-2012)
Excavation Damage (High=65)	Apparent positive trend ✓	*For comparison -- not directly associated with Integrity Management*		✓		



Damage prevention—not integrity management—might explain nearly all the *positive* outcomes for hazardous liquid pipelines.





- Almost two-thirds of the reduction in the total number of hazardous liquid accidents can be attributed to a decline in excavation damage.[35]

- Gas transmission systems have seen more negative trends, generally, in incidents and consequences, but excavation incidents in particular have declined after IM—a change in direction from the previous trend.

- For incidents with death or major injury, excavation damage is the only accident cause with a significant trend since IM implementation. There were six excavation accidents involving death or major injury in 11 years since IM implementation—down considerably from an average of 2/year before IM.

- Injuries due to excavation damage declined by 80% from an average of 3.5/year to 0.7/year.

- The number of high consequence accidents attributed to excavation damage declined from an average of 10/year before IM to 6/year after IM.

- For incidents with fire or explosion, excavation damage is the only accident cause with a significant trend since IM implementation. There were only five excavation damage accidents with fire or explosion in the last 11 years, compared to an average of 2.8/year before IM.

- The amount spilled due to excavation damage declined by two-thirds from an average of 44,000 barrels/year to 14,000 barrels/year.

Damage prevention programs present a much more compelling explanation than IM for reductions in risk over the past decade, particularly on hazardous liquid pipelines. And factoring out damage prevention makes it even more difficult to explain the outcomes we are seeing since IM implementation.

Increasingly more accurate reporting of hazardous liquid accidents in areas that could affect HCAs might help explain some of the differences between HCA and not-HCA outcomes. During the early years of IM implementation, accident reports included an ambiguous reporting element, asking whether the accident was in an HCA. However, much of the hazardous liquid pipeline mileage was not in HCAs but in areas that "could affect" HCAs. So there was a potential disconnect between mileage reported in the annual reports and accidents reported as they occurred.

To the extent that the data became better over time, it would have appeared that the number of accidents in HCAs was rising and the number not in HCAs was declining more than they actually were. This could account for some of the HCA-non-HCA patterns we are seeing in the number of hazardous liquid pipeline accidents, corrosion failures, and spills with environmental consequences since IM implementation. It does not explain the overall outcome trends.

Some general observations on safety outcomes

The weight of the evidence suggests that the expected, high-level safety outcomes from the IM program are not occurring. The twelve external factors I analyzed reflect some likely influences—some positive and some negative. But the overall patterns still suggest no meaningful improvement in hazardous liquid pipelines, and possibly some increase in the risk from gas transmission pipelines.

It is possible that the program was simply over-optimistic in its expectations, and that even negative results have been *less negative* than they would have been without integrity management. In other words, maybe the regulatory baseline (i.e., what was expected in the absence of the rule) was simply inaccurate. The regulatory (cost-benefit) evaluations were in fact less detailed than those for more recent rules. The cost-benefit analysis for the liquid IM rule did not even attempt to quantify the expected results.

But in the rulemaking, PHMSA clearly identified expected reductions in adverse consequences from pipeline accidents and incidents. These expectations generally followed long-term trends, and reflected the logic and promise of "performance-based regulation." They also follow the general logic of today's performance goals, assuming a continual reduction in risk as many programs come together over time to improve the condition and operation of pipeline systems. The program logic and the preambles to the rules set reasonably clear expectations that IM would improve safety in the most important areas while holding onto the safety gains in other areas.

Some have suggested that non-compliance with the IM rules led or contributed to many failures, particularly some higher-consequence incidents. In fact, rulemaking often simply assumes compliance with the rules. But to the extent that non-compliance is might be a significant issue, this would also suggest weaknesses in program design or implementation—gaps in incentives, resources, guidance, or enforcement.

Too early to judge? Several people have suggested that it might be too soon to expect results from the integrity management program, as gas and liquid operators have only recently completed the first cycle of assessments. Implementing new programs nearly always runs into cultural norms, cognitive biases, and often long lead times to ramp up resources, processes, and systems. At the same time, though, this program was designed to “weed out” all of the critical defects and risks incrementally over the 5-7 year assessment cycle. It was front-loaded to require most of the assessments to be done early. Reassessments are aimed at maintaining the status quo. And there is nothing in the design of the program that suggests a clear *mechanism* for getting any better results than what we have now.

My assessment of this: There is probably a lot of truth in the argument that programs often take a long time to effect real change. But the reasons for that are exactly what program evaluation is intended to help sort out and fix. Many of the problems aren’t intrinsic.

This leaves us with a less-than convincing argument that IM has been effective in reducing risk and suggests looking at whether something (or some things) about the design or implementation of integrity management isn’t working as expected. Of course, this could be true even if the expected outcomes were occurring. But the warning flags suggest looking deeper.

Program design

Differences between intended and actual program outcomes can reflect weaknesses in the general logic or design of the program, errors in some of the underlying assumptions, or disconnects between design and implementation. These can be positive or negative in effect.

Performance-based regulation

Integrity Management is one among a broad class of government safety programs commonly referred to as Performance-Based Regulation (PBR). The IM rules require operators to assess, evaluate, repair, and validate—though comprehensive analysis—the integrity of their pipeline segments that could affect high consequence areas. Operators must develop and follow an IM program that provides for continually assessing the integrity of these segments, and take measures to prevent or mitigate the highest threats.[36]

All federal agencies must[37] "to the extent feasible, specify performance objectives, rather than specifying the behavior or manner of compliance that regulated entities must adopt." At the same time, safety regulatory agencies have long recognized that there are many advantages to a performance-based approach. In principle, it puts greater responsibility on companies to manage their own systems safely, with greater flexibility and greater accountability for safe outcomes.[38] This is intended to reduce costs or regulatory burden, provide for faster deployment of new technologies, and encourage companies to go beyond compliance with specifications. It recognizes that things like safety culture—which cannot be regulated—might be more important than simple compliance with rules.

There is a long history of safety regulatory agencies applying performance-based regulation through a set of plans, processes, evaluations, and reviews—drawing on general experience with quality management systems. A fundamental tenet of these approaches is that accidents are not caused by individual human errors or component failures, but by a breakdown in the overall safety management system.[39] A safety management system requires clear and convincing evidence of an alignment between good intentions and real, on-the-ground delivery, and it requires action when that evidence is missing. It generally supports an internally-driven learning culture that aims for continuous improvement.

An ultimate performance-based program would say: *No accidents; you figure out how to make that happen*. No government program has taken PBR that far.[40] All of them fall somewhere on a spectrum of performance-based approaches, sometimes specifying processes or performance characteristics, and usually supplementing more detailed regulatory specifications. All appear to recognize the value of a mix of approaches, depending on the nature of the problem, criticality of the systems, and characteristics of the regulated community.

- *Process Safety Management (PSM)* might be the oldest or most mature PBR – In February 1992, the Occupational Safety and Health Administration (OSHA) in the Department of Labor implemented PSM after several large-scale chemical incidents. PSM is a performance standard requiring employers to provide a comprehensive management program—a holistic approach to integrate technologies, procedures, external standards, and management practices. It includes processes for hazard analysis and audits.

- *International Safety Management (ISM)* – The ISM Code requires that each shipping company have a working safety management system (SMS)—providing safe practices, identifying and managing risks, continuously improving safety skills, and requiring compliance with a wide range of other rules and regulations. It blends prescriptive with performance-based requirements. The ISM Code was adopted by international convention in 1994 and over the next 8-10 years was applied to almost all of the international shipping community. The U.S. Coast Guard (USCG) in the Department of Homeland Security enforces the ISM requirements for U.S. vessels and foreign vessels in U.S. waters.

- *Risk Management Program (RMP)* – EPA's regulations require chemical facility operators to submit risk management plans, which must include analysis of worst-case scenarios, and most of the other common elements of a safety management system. EPA's program adopts all the elements of OSHA's PSM program. EPA's inspection efforts focus on high risk facilities (large population areas, history of accidental releases, and differences in accident rates).

- *Safety Case* – The UK model (also used in Australia) requires companies to submit a safety case for review and acceptance prior to operation. The Safety Case Regulations require a management system to ensure compliance with all relevant statutory provisions, identification of all hazards with the potential to cause a major accident,

The IM rules established 8 elements that must be included in each operator's IM program. At a minimum, each plan must include:

1. *A process for identifying segments that could affect an HCA*
2. *A baseline assessment plan*
3. *An analysis that integrates information about the integrity of the pipeline and the consequences of failure*
4. *Repair criteria*
5. *A process for continual assessment and evaluation*
6. *Identification of preventive and mitigative measures to protect HCAs*
7. *Methods to measure program effectiveness*
8. *A process for review of integrity assessment results*

evaluation of all major risks and implementation of measures to control those risks, audits and reports, and satisfactory management arrangements with contractors and subcontractors.[41]

- *Safety and Environmental Management System (SEMS)* – In 2010, the Bureau of Safety and Environmental Enforcement (BSEE) in the Department of the Interior expanded the agency's approach from hardware-oriented regulation to safety management—blending prescriptive with performance-based requirements in SEMS. The new rule adopted American Petroleum Institute (API) Recommended Practice RP-75. It requires operators to specify how they will manage safety holistically to avoid injury and spills.

- *Safety Management Systems (SMS)* – In 2012, the Secretary of Transportation issued the first Safety Policy for DOT, and subsequently asked each operating administration to develop an SMS implementation plan with regard to its safety oversight of the transportation industry.[42] SMS consists of four main components: safety policy, safety risk management, safety assurance, and safety promotion. The Secretary encouraged operating administrations to promote SMS for each entity it regulates or oversees.

There are several general elements commonly found in PBR approaches for safety programs. Twelve are elements illustrated here. The IM program addresses six of these directly, and partially addresses two others:

1. A safety policy/plan, demonstrating leadership commitment
2. Procedures to identify hazards, evaluate risks, and implement risk controls
3. Operating procedures and safe work practices
4. Pre-startup review
5. Procedures for reporting and investigating accidents and non-conformities
6. Procedures for responding to emergencies
7. Processes for continuous improvement and management of change

	Common elements of Performance-Based Regulations and Safety Management Systems	IM (PHMSA)	PSM (OSHA)	ISM (USCG)	RMP (EPA)	Safety Case (UK)	SEMS (BSEE)	SMS (DOT)
1	A safety policy, demonstrating leadership commitment	(p)	x	x	x	x	x	x
2	Procedures to identify hazards, evaluate risks, and implement risk controls	x	x	x	x	x	x	x
3	Operating procedures and safe work practices		x	x	x	(p)	x	
4	Pre-startup review		x	x	x	x	x	
5	Procedures for reporting and investigating accidents and non-conformities	x	(p)	x	(p)		x	x
6	Procedures for responding to emergencies	x	x	x	x	x	x	x
7	Processes for continuous improvement and management of change	x	x	x	x	(p)	x	x
8	Regular audits, program evaluations, and management reviews	x	x	x	x	x	x	x
9	Defined authorities, responsibilities, and accountabilities, sometimes including stop-work authority	x	x	x	x	x	x	x
10	Training, resources, and competencies to support safety	(p)	x	x	x	x	x	x
11	Employee participation		x	(p)	x	x	x	
12	A strong safety culture—shared values, actions, and behaviors that demonstrate a commitment to safety over competing demands							x

8. Regular audits, performance measurement, program evaluations, and management reviews
9. Defined authorities, responsibilities, and accountabilities, sometimes including stop-work authority
10. Training, resources, and competencies to support safety
11. Employee participation
12. A strong safety culture—shared values, actions, and behaviors that demonstrate a commitment to safety over competing demands

Experience with different PBR approaches has revealed a number of common challenges and some unique challenges that can limit their effectiveness.[43]

- *Performance-based regulation (PBR) brings some inherent challenges*.[44] These regulatory approaches tend to be more data-intensive, and more resource-intensive. They require good quality data, good risk models, and people who can interpret the data effectively. This greater specialization can be particularly challenging for smaller operations, although company size is not necessarily correlated. Some have commented that PBR approaches often don't work because of inadequate resources and expertise. As a result, they can become paper programs with no discernible impact on risk.

> *If you think it's too easy, you're probably not doing it right.*
>
> Linda Daugherty, Deputy Associate Administrator for Pipeline Safety -- at Texas City, TX (Sep. 19, 2012)

- *PBR is also inherently more subjective and judgment-oriented*, and therefore more difficult to enforce. Inspections require an audit approach that is different from more traditional inspection approaches. Many practitioners have pointed out that this requires a different skillset.[45]

- *PBRs are not intended as a substitute for all design and performance specifications.* Specification-oriented rules are still used for critical tasks and where standardization is critical. The IM rules were added to the existing body of pipeline safety regulations.

- *Some PBR's set absolute standards for acceptable risk* (for example, "As Low As Reasonably Achievable" (ALARA)), while others aim to prioritize relative risks. Legislative history can be important in determining which standards are used. IM was designed largely to help prioritize *relative* risk for pipeline systems; repair criteria and some mitigation requirements provide absolute standards for managing risk.

- *The integrity management program for pipeline safety* also brings some particular challenges in data quality, data integration, risk modeling, risk management, metrics, inspection/oversight, and enforcement. These are discussed in other sections of this report.

- *The experience of other agencies with PBR suggests that it needs time to mature*, perhaps more than the 10 years we have observed so far with hazardous liquid pipelines. But the program should stay on the current course only to the extent that the program is conceptually sound and being implemented in a sound way.

The basic logic of performance-based regulation is compelling, but the challenges appear to be equally daunting for the agencies that have tried it. Clearly RSPA (PHMSA) did not identify, consider, and mitigate all these issues in the design and implementation of the program.

Targeting HCAs

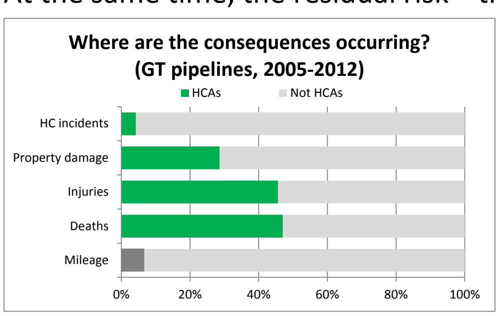
One of the key elements of IM design is the targeting of effort to areas that present higher consequence risks. This rests on two key assumptions: 1) that enough risk is concentrated in HCAs to affect the overall risk and to justify substantial investment of resources there, and 2) that HCA criteria target areas that actually present the highest risk of harm to people and the environment. The evidence supporting these assumptions is mixed.

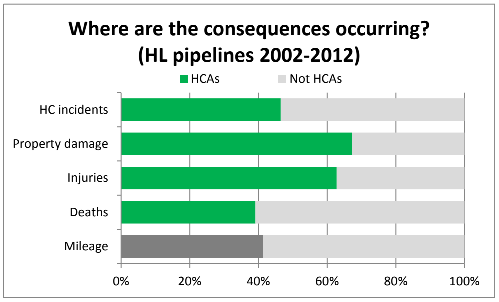
Analysis of the data shows that deaths, injuries, property damage, and the number of high consequence incidents nearly all have occurred at a higher rate (per 1,000 miles of pipe) in HCAs than outside HCAs. The only exceptions are the death rate for liquid pipelines (virtually the same in both areas), and the rate of high consequence incidents for gas transmission pipelines (somewhat higher outside HCAs). These rates suggest that HCA criteria have effectively targeted areas that actually present the highest (or disproportionate) risk.

Consequences - rates per 1,000 miles/year (2002-2012)

		Deaths	Injuries	PD ($millions)	HC incid.
GT	HCA	0.036	0.232	1.871	0.054
	Not HCA	0.003	0.020	0.335	0.087
HL	HCA	0.011	0.039	1.993	0.222
	Not HCA	0.012	0.016	0.681	0.180

At the same time, the residual risk—the adverse consequences remaining outside the scope of the IM rules—is relatively high. For gas transmission pipelines, nearly half of all deaths and injuries, over 2/3 of all property damage, and 96% of all high consequence incidents are still occurring outside HCAs, and therefore beyond the scope of the IM rules. For liquid pipelines, about one third of the injuries and





property damage, 60% of the deaths, and over half of the high consequences incidents are occurring outside HCAs. These numbers suggest some of the natural limitations of a program that applies only to HCAs.

The trends since IM implementation are unclear. I could not detect any meaningful differences in the trends between pipelines in HCAs and pipelines outside HCAs (see the table on page 32). This might be limitation of the data. It might also reflect the broader application of IM assessments and repairs beyond the requirements of the rule. As a practical matter, operators often extend their assessments beyond HCAs as a matter of convenience or interest. The places where they can launch and retrieve inline inspection tools commonly do not coincide with HCA locations, so the assessments cast a wider net. But then the requirements for further investigation, repair, mitigation, monitoring, and reassessment diverge.

The program focus on HCAs was established in law in 2002 (for natural gas transmission pipelines) following implementation of IM for hazardous liquid pipelines. So extensive changes now might require a change in statutory authority.

Assessment tools

The IM rules provide several alternatives for assessing pipeline systems, and in many cases multiple tools are used to detect different kinds of defects.

Pressure tests have long been used as a proof test. Using water in place of gas or hazardous liquid, the system is pressurized to a level greater[46] than the maximum allowable operating pressure (MAOP)[47] to demonstrate the integrity of the pipe with a substantial safety margin. These "hydro" tests are normally done after construction is complete, before the system is put into operation, but they can be used after major repairs as well, or whenever there are serious concerns about the integrity of a system. Hydrostatic pressure tests are required on all new construction now, but systems built before 1971 were exempted ("grandfathered") from the regulations. In a working system, these tests can be expensive, as they shut down operation for the test and require removing all the water from the line after the test. They do not reveal the overall condition of the pipe or any conditions that might be just above the failure pressure of the test. And both time/pressure must be controlled to avoid incidentally weakening the steel.

The gas pipeline that failed at San Bruno, CA in 2010 was grandfathered and had never been hydrostatically pressure tested. It's not clear whether a pressure test at the time of construction would have revealed the defects that ultimately failed many decades later. But many inspectors believe that grandfathering should be ended, and that any pipeline that has never been pressure tested should be tested now. Hydro testing provides a level of confidence in pipeline integrity—a known safety margin—that no other assessment method can provide.[48]

In-line inspection (ILI) tools—called smart "pigs" because of the squealing sound these kinds of tools made as they moved through a pipeline—provide measurements of wall thickness, length and depth of corrosion pitting or excavation gouging, detection of certain kinds of cracks, and measurement of dents or other deformation in the pipe. Different tools typically are used for each kind of threat. ILI tools offer the most extensive characterization of defects and condition of the pipe. But they have several important limitations:

- ILI tools generally can't detect seam cracks or stress corrosion cracking.
- Different kinds of tools (e.g., magnetic flux or ultrasonic) have different strengths and weaknesses in finding different kinds of anomalies, and even within the same type of tool the detection capabilities can vary.
- ILI tools can't detect every defect in the pipe, because of a basic design limitation—most ILI tools advertise a 90% probability of detection, which means that about 10% of defects simply will be missed with a single ILI run.
- For defects that are detected, measurement of their size is subject to a margin of error—typically 10-20% with 95% confidence, meaning that the reported depth and area would be within 10-20% of the actual measured depth and area 95% of the time.[49]
- ILI tools also present a problem of "false calls"—indicators of anomalies that do not in fact meet the detection criteria. False calls are like false alarms; they waste resources, and lead to questioning of results.

How you read these uncertainties can make a big difference. A 90% probability of detection sounds like a lot, but knowing that you are completely missing 10% of the actual defects in the pipe should make people very cautious about how to interpret what they see and the predicted burst pressures that result. There is no way to know what wasn't detected; it's not just the smallest defects, it's a function of whether the pads or gauges on the tool missed a spot. A 10-20% measurement tolerance is another matter; this could be accounted for in the calculations, except that the 95% confidence means 5% of measured values will fall outside the reported range. PHMSA's guidance to operators (FAQ 7.19) requires tool tolerances to be used in the risk evaluation, but provides latitude for the operator to decide how to do this.[50]

Tests comparing ILI tools results and predictions to findings from excavations and actual failures have shown that anomaly depths can exceed the reported depths; anomalies are missed even though their length and width exceed the threshold detection limits of the tools; and *pipe sometimes fails at less than the predicted burst pressure* from ILI data.[51,52]

Direct assessment is a process of integrating information from pipeline characteristics, operating history, and the results of inspection, examination, or evaluation to assess integrity. It includes indirect evaluations of cathodic protection and pipe coatings, and direct examination or testing of certain points in the pipeline to test/verify the reliability of indirect evaluations.

Inspectors have observed that many/most operators have not performed direct assessment correctly; some said they have never seen it done correctly (the first time). And many inspectors and Region Directors suggested that direct assessment should not be accepted as an assessment method under the IM rules.

Pipeline repair criteria

Repair criteria provide the only significant, *absolute* safety standard for pipeline integrity in the IM rules. Without this element, the IM rules govern processes toward prioritizing *relative* risks. The repair criteria put all systems on the same footing with an outcome-oriented specification.

Several inspectors have raised a concern about the repair criteria under the IM rules. In some cases, the rules allow companies to continue operating high pressure systems with known defects, in highly populated areas, with safety margins that might be vanishingly small, based on calculations with many sources of uncertainty.

When pipelines are designed and installed, operators calculate the maximum allowable operating pressure (MAOP) based on the strength and dimensions of the steel, location (for gas systems), and safety factors (typically 1.4x - 2.5x) from the pipeline safety regulations.

Over time, pipelines can lose strength for many different reasons. Corrosion (internal or external) can result in metal loss; excavation or other outside force damage can dent or gouge the pipe; stresses and fatigue can lead to cracks. Part of the reason for the safety factors for new pipe is an acknowledgement that these defects can be introduced and may remain undetected for a long period of time.

The basic safety factor

… used for many decades, provides for operating pressures up to 72% of the specified minimum yield strength (SMYS) of the pipe. SMYS is the pressure at which a pipe can begin to "yield" or deform, and this is typically about 10% below the burst pressure. A safety factor (72% is equivalent to a safety factor of 1.39x since 1/0.72 = 1.39) is intended to keep the pressure below the yield pressure.

For gas transmission systems, additional safety factors are added for higher population areas. A safety factor of 1.39x is used for class 1 locations—the least populated areas. In more heavily populated areas, operating pressure is limited to 40% times SMYS, or a safety factor of 2.50x (1/0.40 = 2.50).

For benchmark purposes, keep in mind that a safety factor of 1.0 has *no* real margin for safety. Operating pressure would equal yield pressure (MAOP = 1.0 times SMYS).

Operators generally assess their pipeline systems through pressure testing, internal inspection tools, or direct (external) assessment. The pipeline safety regulations and industry standards have long included thresholds for repair of pipelines when defects are discovered through these assessments. In general, repairs were to be made back to the original design standard. It was a very high standard, but since there was no requirement to do assessments on any regular schedule the pipe might not be looked at again for a long time.

A relatively large safety margin acknowledged that fact.

The IM rules required (for the first time) baseline and regular assessments in high consequence areas, and it set new standards for repair. While the IM rules were generally intended to add new requirements, not to replace existing standards, some of these new repair standards were in fact more liberal than the old standards, on the assumption that pipelines would be reassessed on a regular basis now and defects would be tracked more closely.

- For liquid pipeline systems, the IM rule at 49 CFR 195.452(h)(4)(B) requires *immediate repair* when metal loss exceeds 80% of the wall thickness or *when a calculation of the remaining strength of the pipe shows a predicted burst pressure less than the established MOP at the location of the anomaly*.

- For gas transmission systems, the IM rule at 49 CFR 192.933(d) requires immediate repair when metal loss exceeds 80% of the wall thickness[53] or *when a calculation of the remaining strength of the pipe shows a predicted failure pressure less than or equal to 1.1 times MAOP at the location of the anomaly*.

- The liquid and gas IM rules apply only to HCAs, and the gas rule does not differentiate by class location in determining the need for or timing of repairs.

The question is whether these requirements provide a sufficient safety factor. So let's look at some scenarios as a pipeline reaches the threshold limits for immediate repair.

The baseline condition: Assume 3 pipelines, all in HCAs, constructed with SMYS = 1500 psi. For a liquid pipeline, MOP must be set no higher than 72% of SMYS (MOP = 0.72 x 1500 = 1080). The same calculation would be used to determine MAOP for a gas pipeline in a class 1 location. In a class 4 location, MAOP would be limited to 40% of SMYS (MAOP = 0.40 x 1500 = 600). The predicted burst pressure would be higher than SMYS, maybe 1650-1750 psi.

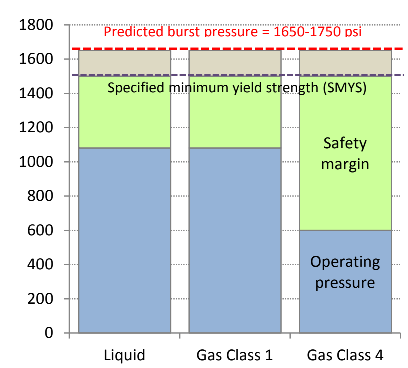


Figure 1: Baseline condition for IM repairs

Then some time passes …

Consider three possible scenarios:

1. **P_{burst} = 1,200 psi**. Assume an ILI tool has detected corrosion in these pipelines with substantial metal loss. If the burst pressure (P_{burst}) is now calculated to be 1200 psi, none of the pipelines would require immediate repair, since P_{burst} > MOP (for liquid) and P_{burst} > 1.1 x MAOP for both gas pipelines.

   The liquid pipeline and the class 1 gas pipeline would be able to continue operating at 90% of burst pressure (1080/1200 = .90), with a safety margin (below the point at which the metal can yield) near zero.

2. **P_{burst} = 1,100 psi**. Now suppose instead that the burst pressure (P_{burst}) is calculated to be 1100 psi. The liquid pipeline does not require an immediate repair (P_{burst} > MOP), the gas pipeline in a class 1 location *requires* an immediate repair (P_{burst} < 1.1 x MAOP), but the gas pipeline in a class 4 location would not (P_{burst} > 1.1 x MAOP).

   The liquid pipeline in this case would be able to continue operating at 98% of the predicted burst pressure, almost certainly higher than the yield strength for this segment, for up to 180 days.[54]

3. **P_{burst} = 670 psi**. Finally, suppose the burst pressure is calculated to be 670 psi. The liquid pipeline and the class 1 gas pipeline would both require immediate repair (670 < MOP for the liquid line and < 1.1 x MAOP for the class 1 line). The class 4 gas pipeline would not require an immediate repair, since 670 is still greater than 1.1 x MAOP (1.1 x 600 = 660).

   Now the gas pipeline in a class 4 location would be able to operate at 90% of burst pressure, with a safety margin relative to yield strength of near zero)—compared to a required safety factor of over 2.5 (40% x SMYS) when it was originally constructed.

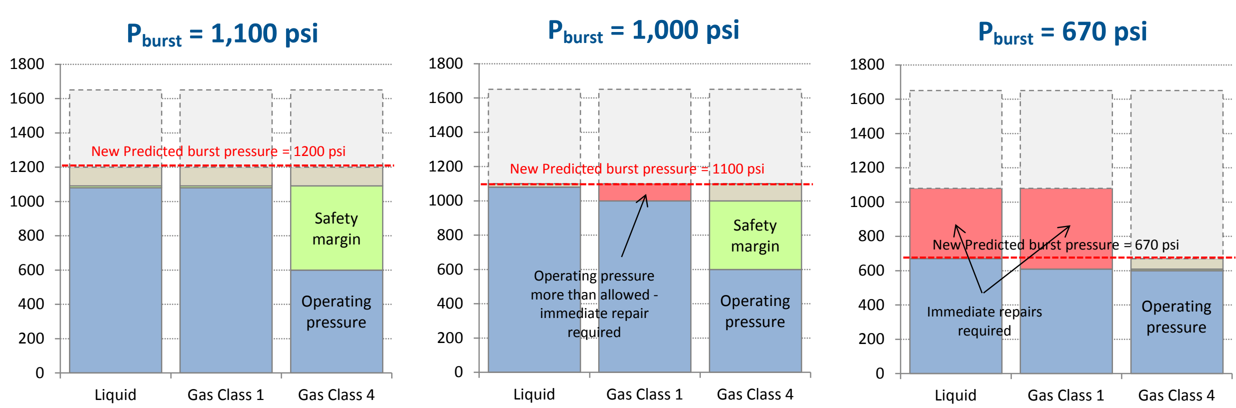


Figure 2: Three scenarios for IM repairs

In the first and third cases above, the pipelines would be permitted to operate in a range very close to their burst pressure, and even closer to the point at which the metal would yield under pressure at the point of the defect. And *all of this assumes that the operators' calculations are correct*, even though there are many uncertainties in the data that are much larger than the safety margins—a probability of detection of 90% or so (the advertised PoD by most ILI vendors), and ILI measurement accuracy of ±20% with 95% probability (generally advertised as well).

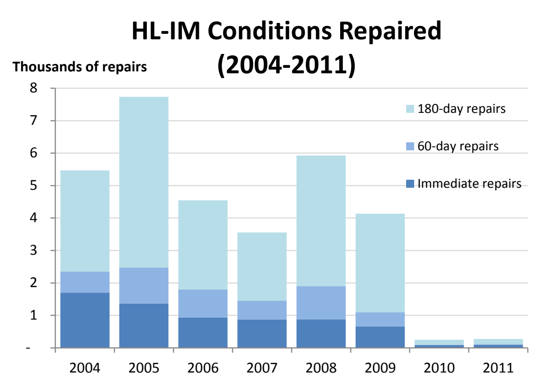
Given these uncertainties, *calculated* safety factors ranging from 1.02 to 1.1 are likely to be overly optimistic and not very reassuring, particularly in HCAs. Some inspectors (experienced professional engineers) have described this situation as "dangerous." Several others in the program agree that the standard is probably too low. And research testing has shown that pipe does in fact sometimes fail at less than the calculated burst pressures.

It also appears that the repair criteria, by linking the criteria to burst pressure instead of MAOP, and not distinguishing class location for gas pipelines, means that *high population areas are afforded no more margin for safety than low population areas*—a significant conceptual departure from the basic pipeline code. Given the high cost to replace a pipeline, this suggests that pipelines in more populated areas (class 3 and 4 locations) probably have not been repaired as often as class 1 locations. PHMSA does not collect data to determine whether this is in fact happening. That's an important gap by itself.

Reassessment intervals

The time between assessments is logically intertwined with the repair criteria, as the basic idea is to repair time-dependent threats that could grow to a critical size by the next reassessment, and to reassess before that growth could occur. The rules required operators to set reassessment intervals based on the risk factors specific to its pipeline, but the intervals were not to exceed five years (for liquid pipelines) or seven years (for gas pipelines) unless the operator provided to PHMSA an engineering basis for a longer interval.

If reassessment intervals are set sufficiently short (so that time-dependent threats can't grow to critical size, given the repair criteria), we should see a significant reduction in immediate repair conditions as reassessments are started. In fact, this is what we see in the data for liquid pipeline operators as baseline assessments were completed in 2008-2009 and reassessments followed.



The evidence here suggests several things (for liquid pipelines):

- The reassessment intervals are sufficiently short to work effectively. This, combined with any followup monitoring operators are doing, appears to provide protection against growth of the kinds of defects that are targeted in the assessments.
- Reassessment intervals, particularly for liquid pipelines, might be too short. If these intervals are too conservative, the safety the program is getting could be at a cost that is too high. Analyzing this would require data beyond what we have, so it is beyond the scope of this evaluation. Further research might look at the actual growth of defects between successive assessments, across the industry.
- Repairs from the baseline assessments through 2009 evidently cleared out a lot of defects within the detection limits of the tools that were used.
- Operators might not be taking advantage of the most liberal repair criteria. While some inspectors have flagged warnings because of the reduced safety factors allowed in the IM rules, liquid pipeline operators evidently are matching the actual repair criteria they use to the reassessment intervals. This doesn't mean the standard is right, of course.
- The repair rates—particularly immediate repairs per 1,000 miles of pipeline assessed—cannot be determined reliably[55] from the data we have, although this is an obvious indicator of whether operators targeted their highest risk pipe for assessment first. But given the number of liquid pipeline miles assessed in 2004 (about 23,000 miles) and in 2010-2011 (24,000-25,000, respectively), it is reasonable to assume that the miles assessed each year has remained fairly constant as operators balance/spread their costs out over the years. This suggests that the rate has changed in proportion to the numbers in the graphic on the previous page.
- The defect detection rates in 2010-2011 do not help explain the safety outcomes we have seen. The big decline in these rates (for liquid systems) is in sharp contrast with the increase in accidents, property damage, corrosion and material failure, product spilled, and spills with environmental consequences—particularly in HCAs. In fact, these opposing trends make the safety outcomes all the more puzzling.

Data for gas transmission systems are inconclusive. Here, too, the data were not collected in a way that permits easy evaluation of what operators were finding and fixing. Immediate repairs per 1,000 miles assessed appear to be about 60 percent lower for reassessments than for baseline assessments—compared to a drop of about 90 percent for liquid pipeline—but the data are much more limited for gas transmission systems.

Testing the assumptions in program design

The program design assumed that …	This assumption turned out to be …
HCA criteria target areas that are actually highest risk re: harm to people and the environment	*Substantiated by experience* – For most serious consequences, the actual rates per mile have been higher in HCAs than outside HCAs (subject to some data limitations).
Enough of the overall risk is concentrated in HCAs to affect the overall risk and to justify substantial investment of resources in the IM program	*Partly true, but weak* – Focusing on HCAs cannot affect the top-level numbers enough, especially for gas transmission systems where most of the deaths and injuries have occurred outside HCAs.
Assessment technology is sufficient to detect and characterize defects or anomalies accurately	*Largely true, and maybe overrated.* Operators have found and fixed tens of thousands of defects. But risk models generally do not account for uncertainties re: probability of detection or tool tolerances.
A 5-year reassessment interval is sufficient to detect progressive deterioration before a pipeline fails, and is a reasonable length of time to re-check other hazards or threats to the system that might develop	*Substantiated by experience*. The number and severity of defects detected have dropped considerably as reassessments have begun—as expected.
Repair criteria are complete and targeted to the actual highest risks, and will arrest progressive deterioration before failure or the next assessment	*Seriously overestimated*. Repair criteria permit the operation of high pressure pipelines in high population areas with known defects and almost no safety margin.
Operators' skills and resources are adequate to identify systems that could affect HCAs, conduct assessments, interpret and integrate data, evaluate risks, and mitigate the most serious risks	*Probably true in some areas, much less so in others* … discussed in more detail in Program implementation section of this report.
A system-wide approach is a more effective, and in most cases, more efficient means of inspecting pipeline systems and evaluating pipeline integrity compared to inspecting small segments of systems	*Not demonstrated*. The data do not show the expected safety outcomes occurring generally.
One size does not fit all …the increased flexibility that comes with a performance-based regulation should permit adaptation to fit the unique conditions in each pipeline system and encourage development and use of new technologies	*Demonstrated, but the outcomes are not so clear* – the data do not provide any clear evidence that the expected safety outcomes are occurring. But it is clear that inline inspection tool technology has advanced substantially.
PHMSA's inspections would require a change in approach—from inspecting for compliance with specifications to auditing processes.	*A longer process than anticipated* … discussed in more detail in the Inspection/oversight section of this report.

Some general observations on program design

Performance-based regulation is an appealing framework for improving safety. It appears to be logically sound and conceptually efficient. PHMSA (RSPA at the time) developed the IM program aware of many pitfalls with PBR, drawing from other agencies' experience, and designing the program in some ways to try to minimize these pitfalls. But well-tested models in this area are not easy to find. Agencies are continuing to explore many different variations.

One of the key design features in the IM program is a focus on pipelines that are in, or could affect, an HCA. There is compelling logic for this focus as well, and the data show some high numbers and rates of serious consequences in HCAs. But the data also show that a focus on HCAs is limiting—probably more so than expected in the rulemaking process. More than half of the really bad safety outcomes still occur outside HCAs and beyond the regulatory reach of IM.

Many inspectors and program staff have observed that the IM program was developed from industry standards, not the agency's own analysis or a logical review of the safety benefits or costs. Assessment methods, repair criteria, reassessment intervals, risk modeling approaches, performance measures, even the estimated costs of various options—all were drawn from the regulated industry and a growing body of industry standards at the time.

Some inspectors have suggested that grandfathering older pipe (which is currently exempt from the hydrostatic test requirements) is a big gap. The Pacific Gas and Electric pipeline that failed catastrophically in San Bruno, CA was grandfathered. While the issue of grandfathering was decided outside the IM rules, the more general issue of testing or examination is an important element of the IM program. Several inspectors have suggested that the agency should not allow direct assessment when companies have limited records about the pipe they have in the ground. One inspector observed that he has never seen a company perform direct assessment correctly the first time; many were very skeptical of direct assessment.

Inspectors commented extensively on what companies are actually doing to implement the IM program, where they seem to be struggling or doing particularly well, the challenges in inspection/oversight and enforcement, data quality, data integration, risk models, inspection protocols and the inspection process, metrics, and the idea of setting performance goals for companies. These are all addressed in subsequent sections of this evaluation report.

Given the overall objectives of IM and program logic, the success of the program is tied to the basic design assumptions. Where they are not well supported by data or analysis, it shouldn't be surprising to find particular challenges that become more evident in IM implementation.

Program implementation

There are nearly always disconnects between how a program was designed and how it was implemented. This is not always a bad thing. Gaps between design and implementation can result from miscommunication or competing goals, but they can also reflect adaptation to circumstances that were not anticipated in the original design. The standard should not be whether someone did what they were told or expected to do, or even what they said they would do. Ultimately it's about whether the implementation works. And when there is a difference between design and implementation, we need to understand why.

The theory of change

Programs are implemented in a social and organizational context. In evaluating program design and its implementation, it can be useful to look at the theory of change—understanding the drivers (psychological, economic, sociological, physical processes) by which change comes about, and designing strategies and incentives to help ensure that the program is implemented as designed. For IM, the theory of change implicit in the rules and implementation plans seems pretty simple—it's more of a linear engineering model than an organizational behavior model:

> The agency would outline the processes …
> → companies would implement them,
> → the agency would inspect companies' processes to monitor implementation, and
> → safety would improve as result.

There are many questions that might be asked as a program like this is being implemented; for example:

- Will companies add resources for IM or take away from what they had been doing?
- If no new resources, what might they stop doing or scale back?
- How do those activities relate to the IM activities?
- What incentives will drive decisions on assessment and maintenance?
- Do companies have the expertise to implement IM?
- If not, how will they get it?
- If they contract out for evaluative work, how will they develop a better understanding of their systems (one of the fundamental principles of IM)?
- How will companies adapt when they run into things the program didn't anticipate in the design?
- How will we know if the program is working (or not)?

- What kinds of unintended consequences might we anticipate?
- How would we detect them?
- What kinds of external factors could influence what companies do?

PHMSA used the monitoring process to make adjustments. Through two rounds of IM inspections and "reset" meetings, inspectors and program managers observed implementation of the IM processes, evaluated discrepancies between design and implementation, and adapted guidance to the industry to mitigate these discrepancies. An extensive set of frequently asked questions (FAQs) grew to supplement the regulations. At the same time, the inspection protocols grew to provide inspectors with better guidance on how to evaluate an operator's IM processes. But there was *no systematic evaluation of the psychological, economic, sociological, and organizational processes* underlying what companies were actually doing.

In practice, some companies added resources to implement IM and many did not.[56] We don't really know what activities these companies stopped doing or scaled back. Some companies deferred assessments and other work at certain times of the year based simply on budgetary factors. Many companies contracted out a substantial fraction of their IM assessment, data interpretation, analysis, and risk modeling; and many of these companies did not develop the technical expertise to evaluate what the contractors provided.[57] These kinds of issues can affect successful implementation of a program, but (except, perhaps, by chance) the implementation team had no expertise in organizational behavior to evaluate them.

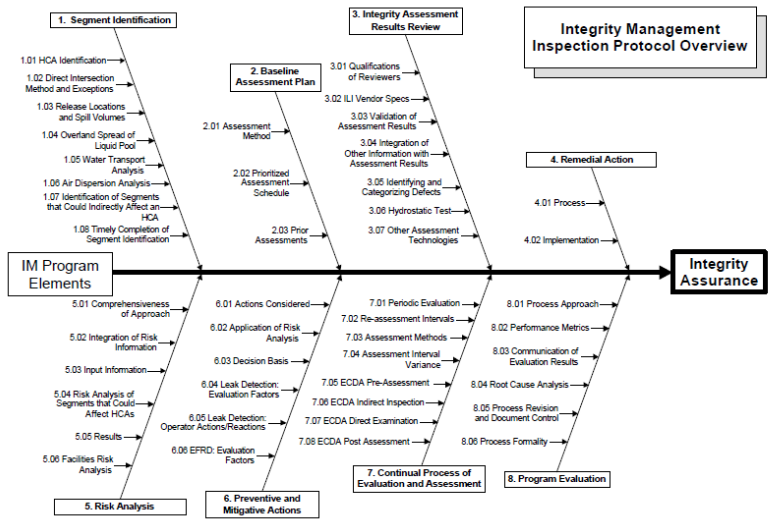
This gap is all the more conspicuous in a performance-based approach to safety regulation, where the focus is on leadership, processes, continuous improvement, management of change, program evaluation, management reviews, authorities and accountability, and organizational culture (see the 12 common elements of PBRs). *These are not engineering specialties (every inspector is an engineer). Evaluating them requires expertise in organizational behavior, which the agency does not have.*[58]

Managing implementation

The IM rules established eight elements that must be included in each operator's IM program—a detailed plan to address eight more-or-less sequential program elements. And during the first two rounds of inspections, the agency focused heavily on evaluating these areas:

1. A process for identifying segments that could affect an HCA
2. A baseline assessment plan
3. An analysis that integrates information about the integrity of the pipeline and the consequences of failure

4. Repair criteria
5. A process for continual assessment and evaluation
6. Identification of preventive and mitigative measures to protect HCAs
7. Methods to measure program effectiveness
8. A process for review of integrity assessment results



The first two rounds of inspections under the IM program were focused on helping operators develop a basic program, letting them grow with the program and expecting them to make it better over time.[59]

Pigging and digging and targeting the highest risks first

Pigging (running in-line inspection tools) and digging (excavating segments of pipe with more serious anomalies to confirm the tool measurements and make repairs when needed) appear to be staple activities under the IM program. Field inspectors nearly all noted the detection and repair of sometimes hundreds of defects on systems that would not have been detected without the IM program. Detection and repair of defects is an important intermediate outcome for the program. The general reasoning is that defects that are detected and repaired are potential/future accidents that won't happen; and that as more of these defects are repaired over time, systems should be in better condition.

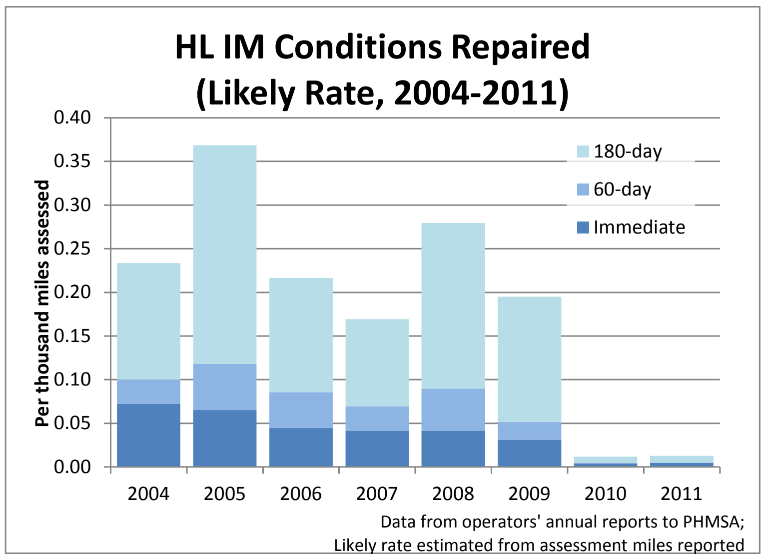
Inspectors have also reported that the technology of in-line inspection (ILI) tools has clearly advanced to detect defects that would not have been detected before, and they have noted that these advances in technology were driven by the IM program requirements. They are seeing much more use of the tools than before IM. And in some areas, they are noticing fewer big failures or accidents.

As more defects are repaired over time, systems should be in better condition. So the program expected *a declining number* of defects over time, particularly in HCAs, as operators find and fix defects. If operators were targeting the highest-risk segments of their pipeline systems for assessment, we should have seen *a decline in the rate* of defects even over the baseline assessment period. Unfortunately, the program didn't collect the right data to tell us whether this was happening or not; but the evidence suggests the rate probably declined as expected …

- The number of immediate repair conditions declined fairly steadily from 2004-2009 then dropped by an order of magnitude.

- The denominator (to calculate a rate) is much more ambiguous. The number of "baseline miles completed" dropped even more than the number of immediate repair conditions, indicating that the rate actually increased. But this isn't the whole picture.

- The IM rule required operators to assess at least 50% of their HCA pipe by September 30, 2004. Since only 23,000 miles of pipe (about a third of the total HCA mileage at the time) was reported completed in 2004, a significant fraction must have been done in 2002-2003 or even earlier. The rule permitted assessments conducted after January 1, 1996 to be credited as a baseline assessment. This meant that 5-year reassessments were beginning in 2007, or even as early as 2001 (there was no place to report this in annual reports until 2010). But since the repair criteria were not in effect before the IM rule, some reassessments were in effect "first-time" IM assessments.



- The most likely scenario is that many/most operators were spreading their assessment effort over time to match resources.

- First-time IM assessments—any assessment conducted after the IM rules were effective—probably continued at a fairly steady level through about 2008. This would include baseline assessments but also a growing number of reassessments for pipe that was credited with a pre-IM baseline assessment. If so, the trend in the rate would match the trend in the number of defects found.

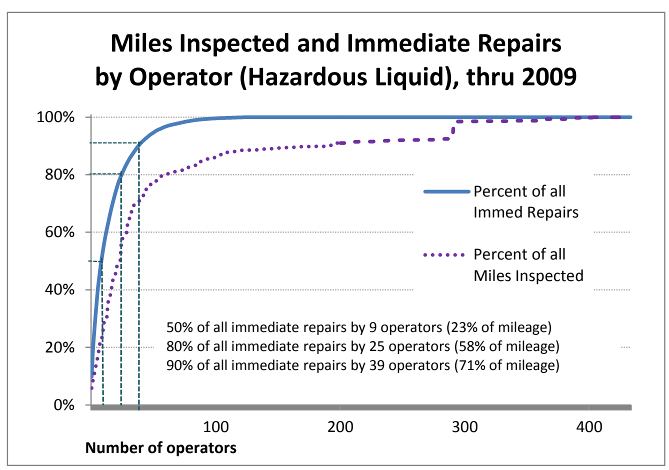
Liquid pipeline operators reported over 6,000 immediate repair conditions; 4,000 60-day repair conditions; 20,000 180-day conditions; and nearly 70,000 other conditions discovered and repaired over the six-year period 2004-2009—an average of about 24,000 total each year. If defects that are detected and repaired are potential accidents that won't happen, that's a lot of risk reduction. But it's also hard to reconcile that conclusion with the accident data. There are only about 120-150 "significant" liquid pipeline accidents each year over the entire system. And those numbers haven't declined very much.

Some possible explanations: 1) companies were doing these kinds of repairs already before the IM requirements (i.e., *nothing new here*); 2) the thresholds companies are using for making repairs are too conservative (i.e., *doing too much*); 3) defects were clustered, such that many were critical where only one would ultimately fail (i.e., *the problems are deeper than each individual defect*); or 4) these defects signaled a critical deterioration in the overall pipeline infrastructure that was caught just in time.

In fact, we know that some companies were doing these kinds of repairs before IM. The preamble to the liquid IM rule estimated that about 10% of operators already had functional IM programs. But by the time operators reported their assessments completed, it was obvious that the pre-IM rate was about 50% of the post-IM rate of assessments. And companies made repairs based on other findings (leaks, etc.). So this is probably part of the overall explanation.

It is possible that more repairs were done than needed to be done, or at least that the conditions were not as "immediate" as indicated. This is not an easy argument to make, though, given the previous discussion of repair criteria and vanishing safety margins. Did all of the reported repairs actually rise to the level of the thresholds in regulation? We don't know, because the program doesn't ask for this data.

Some defects probably were clustered. Annual report data through 2009 show that 50% of all immediate repairs were done by 9 operators, 80% were done by 25 operators, and 90% were done by 39 operators. These same operators managed a disproportionate amount of pipeline mileage at the same time, and these immediate repairs tended to be spread over multiple years. But there is still some significant clustering of the data.



The argument that operators caught these defects just in time is also reflected in many inspectors' argument that we don't know what would have happened without integrity management. There might be some truth to this argument too, but the long-term trends and the analyses of pipeline age/vintage suggest otherwise.

Going beyond pigging and digging

Pipeline operators have run lots of pigs and made thousands of repairs, and these benefits have gone beyond the HCAs where assessment is required. Many companies have also invested in making more of their systems "piggable" in order to get the benefits of detailed, internal characterization of their pipelines. But inspectors and investigators have observed during inspections and after pipeline ruptures that companies often have not addressed known problems.

Most inspectors observed that a lot of companies are focused on pigging and digging.[60] "The assessment rules were pretty specific, black and white; the rest was squishier, so it was harder to sell within a company." The annual reports reinforce this as most important; that's what we're asking about, and companies are doing what the program is measuring. Inspectors don't see most companies seriously implementing other IM requirements like data analysis, data integration, prevention and mitigation measures, leak detection, metrics, updating risk models, and generally learning about their systems.

There are some issues even in the use of in-line inspection (ILI) tools. Inspectors have observed that the tools (or the analysis of tool results) often miss things (like cracks) that later cause a leak or rupture, and that crack-detection tools are sometimes misused. The tools might not be good enough to detect stress corrosion cracking or selective interference. And sometimes operators might be "missing the forest for the trees"—focusing at the anomaly level, not the system level—even though the intent of the IM program was to get the big picture.

To some extent, this may be simply a resource issue. Some companies increased their resources to comply with IM and put a lot of effort into it. Many companies didn't hire more people or increase their budgets to implement the program,[61] so their overall effectiveness depends on where they "harvested" resources. And some companies have deferred maintenance or repairs based on timing in the budget year. Some companies—especially larger ones—might have their own specialists who have more expertise than the tool vendors, and more consistent policies. Smaller companies have tended to rely more on contractors to help assess the risks in their systems. This can provide more specialized expertise, but doesn't necessarily leave a company with a broader understanding of its own system.

All of this gets at one of the basic design assumptions for the IM program—that *operators' skills and resources* are adequate to implement the program. The evidence suggests this is very uneven throughout the industry, maybe much less true than people anticipated.

Safety culture

Over the last 25 years or so—since the explosion of a nuclear reactor at Chernobyl, and increasingly over the past decade—people have come to recognize the importance of safety culture in managing high-risk systems. Safety Culture is defined (by DOT) as the shared values, actions, and behaviors that demonstrate a commitment to safety over competing goals and demands.[62] It reflects whether an organization seeks to be a model for good safety practices or whether it is simply content to keep one step ahead of the regulators.[63]

Safety culture clearly shares many common elements with safety management systems and performance-based regulatory approaches to safety (see sidebar)—including the importance of leadership, employee participation, continuous learning, reporting systems, accountability, and training and resources. But a focus on safety culture is generally missing as an explicit element of these systems, as it is for IM.

Inspectors' observations on this are mixed. Some note that IM has been a learning experience for everyone, even if it is still at an early stage after 10 years. Operators are (sometimes) changing their programs when things go wrong, and some are managing change better than expected. But many other comments from the field reveal vulnerabilities tied to safety culture:

- *Companies are focused more on building than putting dollars into maintaining existing assets.*
- *It's sometimes challenging to make repairs in HCAs (pipe under rivers, etc.) and companies are constantly explaining why they can't make repairs. But it's ultimately about cost.*
- *We should be looking at [company incentive] structures. There is a huge incentive not to spend money on repairs.*
- *Incentive provisions in contracts emphasize timeliness, not safety.*

Some of the key elements of a strong safety culture:

1. Leadership is clearly committed to safety;
2. There is open and effective communication across the organization;
3. Employees feel personally responsible for safety;
4. The organization practices continuous learning;
5. There is a safety conscious work environment;
6. Reporting systems are clearly defined and non-punitive;
7. Decisions demonstrate that safety is prioritized over competing demands;
8. Mutual trust is fostered between employees and the organization;
9. The organization is fair and consistent in responding to safety concerns; and
10. Training and resources are available to support safety.

Drawn from a research paper prepared for the U.S. DOT Safety Council, May 2011

- *In some cases it's obvious that operations and maintenance are overly constrained by funding.*
- *Some companies have the data to make management decisions; others don't have the resources or management support [for good quality data].*
- *Management commitment is key—the root cause of failures isn't a lack of engineering expertise.*

This last point is worth highlighting. Weaknesses in a company's safety culture are not individual failures or personal fault. These are *organizational* failures. There is no data to demonstrate the extent to which operators have strong or weak safety cultures. The evidence is all anecdotal. But if safety culture is important, that is another gap by itself.

There is one area where program guidance to operators could undermine safety culture. This is how companies define "risks." One of the Frequently Asked Questions (FAQs) posted on the agency's website asks: *Can operators include potential business consequences (e.g., curtailments, plant shutdown) in its risk determinations?*[64] The answer emphasizes safety and environmental risk, but clearly supports including other kinds of risk as well.

> The focus of the integrity management rule is reducing the risk of pipeline failures to high consequence areas. The integrity management programs developed to comply with rule requirements should include the use of risk analysis to support operator integrity decisions. Operator risk analysis processes require the evaluation and measurement of both the probability and consequences of pipeline failures. *The appropriate consequences to be included in these risk analyses depend on the decisions that are being supported by the risk analysis results*. [italics added]
>
> In the context of fulfilling requirements of the integrity management rule, operators should maintain a focus on the risk of failures to high consequence areas. Consequently, operators should emphasize those consequences that are considered in the definition of high consequence areas (i.e., human health and safety, environmental protection, property damage, local economic impacts).
>
> *If consequences considered in the risk analysis are expanded to include consequences related to operator business performance*, then the operator must provide assurance that this approach does not skew decisions away from protection of HCAs …

This approach is consistent with the Federal standards for regulatory evaluations, which also must consider *all* impacts of a rule. Operations and maintenance decisions will always be based on many factors. But the guidance might tend to undermine the safety and environmental focus of the program, and a commitment to safety over competing goals and demands, by permitting things like delivery schedules to be factored into the results of a risk analysis. It potentially deprives decision makers of clear, separate estimates of *safety* risk to use in making decisions. Those tradeoffs can be buried in the model.

Inspectors have pointed out that they don't actually allow operators to include business (or profitability) factors in their risk models, and that if there was any ambiguity in the FAQ they took care of it in the inspection process. That is helpful, but it is a tenuous position to rely on inspectors' institutional memories to correct for known ambiguities in the guidance to operators.

The question itself reflects another phenomenon—*risk homeostasis*—that safety regulators have observed for a long time—safety features and benefits are often converted into production gains. For example, better quality pipe led to higher operating pressures (the MAOP rule). In the IM rule, more frequent assessments led to lower safety margins in repair criteria.

The FAQ effectively adds non-safety risks as legitimate factors to be considered in a risk analysis that was intended to achieve safety outcomes. If, and to the extent that, operators are incorporating financial and economic risks into their risk assessments, this factor could help explain the gap between the expected and actual safety outcomes from the IM program.

Adding risk into the system?

In trying to find an explanation for the safety trends we're seeing, one inspector suggested that maybe IM itself is doing something to *add* risk into the system. That seems a strange idea at first, and one inspector's speculation isn't evidence, but there is some considerable support for this idea in the literature on risk and the experience of accident investigators.

There is an inherent risk in systems that are both complex and tightly-coupled, which can lead to what are commonly known as "*normal accidents*."[65] Natural gas and hazardous liquid pipelines seem to meet these criteria. These are complicated systems, in difficult and sometimes hostile environments, requiring designs with many interactions that often are not visible. Most of the pipe is buried. Since no operation is perfect, there will be failures. And the potential for a system accident with high consequences can increase in a poorly-run organization, as there are more possible failures to interact in unexpected ways.

Maintenance can seriously damage a system.[66] Accident investigations have shown repeatedly that maintenance lapses can create latent (hidden and dormant) conditions for failure. Any time a system is disassembled and reassembled, or disturbed in any way, the potential for failure increases. Pipeline repairs add new materials, and the data show that newer pipe is the riskiest pipe. This is partly construction defects, and partly the activity and potential for damage when working around the pipe. Inspectors note that maintenance issues have already been brought up following several pipeline accidents, and in some cases operators have adjusted maintenance practices as a result.

Preventive and mitigative measures could create new vulnerabilities, depending on how they are designed and implemented. For example, the addition of automation to open and close valves is a protective mechanism to mitigate the effect of a release. But the addition of emergency flow restricting devices can, and sometimes do, cause unintended consequences such as inadvertent closures.

One of the effects of continually tightening up safety practices is to increase the likelihood of violations being committed—a process of trying to adapt to procedural over-specification.[67] Redundant backups increase the interactive complexity of a system, and can increase the probability of unforeseeable common-mode failures.[68] They can also make the system more opaque to the people who control it. Control rooms similarly make a system more complex and opaque, allowing the buildup of latent conditions.

These illustrations, of course, are not to suggest that companies should not do repairs, maintenance, or preventive and mitigative actions. They suggest a need for continuing awareness of the implications.

Some general observations on program implementation

The more basic and familiar requirements of IM—developing plans, performing assessments, making repairs based on those assessments—it seems nearly everyone agrees are happening. Improving the quality and use of data, developing preventive and mitigative measures, evaluating program effectiveness, managing change—have proven to be more difficult to implement. What distinguishes these activities is that they are relatively newer and less familiar. They are also associated with specialized areas of expertise that might not be obvious, and largely seems absent in IM implementation.

Safety culture is a concept that has been around at least 25 years. But only recently have regulatory agencies begun trying explicitly to incorporate it into their regulatory frameworks. NRC and BSEE, in particular, have developed and published policy statements on safety culture. But there is no single, common definition of safety culture and there is an array of overlapping elements deemed important in establishing a strong safety culture. Many of the regulated industries (including especially the oil and gas industry) deal with at least a half dozen different safety oversight agencies. If safety culture is important, it is probably important to address it across all federal agencies. And additional research might be needed to measure it consistently and to understand how agencies' actions affect it—both positively and negatively.

Risk modeling and risk assessment

Studies have shown that the two characteristics most likely to distinguish safe organizations from less safe ones are, firstly, top level commitment and secondly, the possession of an adequate safety information system.

- James Reason, in Managing the Risks of Organizational Accidents

Risk models are the heart of an operator's risk management system. Models combine data and other information into a form that highlights differences in risk. They provide a way to make sense of sometimes-vast amounts of information, narrowing the focus to factors that affect the probability or consequences of a failure. Risk models can be used to:

- Help understand how the system works, how the parts fit together, and how the system might fail
- Prioritize assessments, based on relative risk, to check the condition of the system
- Set intervals for reassessments and for periodic risk evaluations
- Target risk evaluations, based on the risk factors that are most important or influential in the results from the model
- Help identify and diagnose risk issues that might be hidden in the data
- Prioritize and focus mitigation actions
- Justify additional resources for risk management

Risk models used for risk assessment

Under the IM rules, each operator must establish an integrity assessment schedule that prioritizes pipeline segments for assessment, and set reassessment intervals.[69] These schedules must be based on all risk factors that reflect the risk conditions on the pipeline segment. The regulations identify at least 18 risk factors that must be considered, and an appendix provides further guidance on evaluating risk factors. The rules also require generally that operators must follow recognized industry practices unless otherwise specified or an alternative practice is supported by a reliable engineering evaluation.[70]

For liquid pipeline systems, API standard 1160 defines risk as a function of the likelihood and the consequences of a release. It does not present any specific methods for evaluating risk, but discusses several common elements in the variety of methods:

- Identifying potential events or conditions that might present a risk to system integrity

- Collecting and analyzing data to determine the likelihood and consequences of release for each of these events or conditions
- Ranking the results
- Identifying and evaluating options to mitigate the risks
- Providing feedback based on maintenance findings
- Reassessing the risks

The API standard addresses data quality, dealing with uncertainty, resource requirements, feedback from actual field measurements, and review of the data and model results by knowledgeable people—to ensure that the results "make sense." It suggests that risk assessments should be investigative in nature, and predictive of future risk.

The API standard also provides many examples of risk factors an operator might use to assess likelihood and consequences in a risk model. It discusses generally the process of converting these factors into scores and combining the scores into a risk value—illustrated by summing the product of the likelihood and consequences across all the failure modes considered in the model.

For gas transmission pipeline systems, ASME standard B31.8S describes risk as the sum of probability times consequences (P x C) for each of nine threat categories. It presents four methods for operators to use in risk assessment:

- *Subject matter experts*, combined with information from the technical literature, to provide estimates of the likelihood and consequences of failure for each threat. This kind of analysis can be fairly simple, with values ranging from 1 to 3 to reflect high, medium, and low probabilities or consequences. It provides a more subjective assessment than the other methods.

- *Relative risk ranking models* – using more extensive data and system-specific experience – to score, weight, and sum the weighted scores over for each of the major threats and consequences, providing a relative risk value of each segment.

- *Scenario-based models* – including event trees, decision trees, or fault trees – develop risk values for events of series of events.

- *Probabilistic models* – a more complex and data-intensive approach to calculate absolute probabilities and consequences at the component level, providing output that can be compared to acceptable levels of risk established by the operator.

The ASME standard addresses data quality (for example, using default values for missing or questionable data), and requires validation (using the actual results from inspections or examinations to determine if the models are correctly characterizing the risks).

The Pipeline Risk Management Manual[71] describes three general types of models—matrix models, probabilistic models, and indexing models (which the author recommends). Most of the book elaborates on the indexing model (corresponding closely to the relative risk ranking model in ASME B31.8S). Muhlbauer suggests that, if the need is to weigh immediate risk trade-offs or perform inexpensive overall assessments, indexing models might be the best choice. He notes that these kinds of scoring systems are common in many applications beyond pipeline safety.[72]

- Muhlbauer's risk evaluation framework derives a relative risk score from the sum of probabilities for each of four equally-weighted threat categories (3rd party damage, corrosion, design issues, and incorrect operations), divided by a leak impact (consequences) factor—the product of four additional values. Each threat category consists of 5-21 values, each value weighted from 2-35 points, adding to 100 points in each category.

In an appendix, Muhlbauer also illustrates two specific models, chosen to represent the many systems developed by consultants and pipeline companies themselves. Each of these computes probability and consequences in a different way.

- Model 1 derives a relative risk score from the sum of six equally-weighted factors contributing to probability (soil & joining methods, corrosion, damage, hydrostatic test history, leak/rupture history, and pipe condition), multiplied by a consequence formula with six weighted factors.

- Model 2 derives a relative risk score from the average of probabilities for each of seven equally-weighted failure modes, multiplied by the average of five weighted consequence factors (life, property, loss of service, failure cost, and environmental effects). Values include more qualitative information. Weighting for the consequence factors is determined by each operator.

The appendix to the IM rule also uses a relative-risk indexing model to illustrate the risk evaluation required by the rule.

- The guidance derives a risk score from the sum of 18 risk factors, most of which are related to probability (e.g., pipe wall thickness, known corrosion, results of previous testing), some related to consequences (e.g., location near population or environmentally sensitive areas, ability to detect and respond to a leak). Values are scaled or indexed (1 is low, 3 is moderate, 5 is high); and tables are provided to illustrate values for leak history, pipeline diameter, product risk, and age.

Inspectors report that most or nearly all pipeline operators have used, and continue to use, an *index scoring model* to prioritize their integrity assessments.[73]

Notice both the similarities and differences across these models. Summing and averaging are mathematically equivalent, and multiplying and dividing are mathematically equivalent; multiplying then adding is *not equivalent* to adding then multiplying. Most of the models (except the appendix to the IM rule) use a general concept of multiplying probability times consequences to estimate risk—a very standard formulation—but use different factors, different weights, and combine probability and consequence factors in different ways.

Does it matter? Well, yes … it matters a lot. The differences reflect a fundamental lack of scientific underpinning in this kind of model. And even the similarities conceal some deeply flawed assumptions and systematic errors.

Index-scoring risk models may be undermining integrity management

In the report of investigation on the San Bruno pipeline explosion, NTSB raised three concerns with the risk models pipeline operators are using:

- the quality and completeness of the records that are used,
- the extent to which operators are incorporating leak, failure, and incident data in evaluation of their risk models, and
- the weighting of risk factors

There are, in fact, many more serious, documented issues with index scoring risk models[74] beyond these. Consider the risk factors and modeling illustrated in the Hazardous Liquid IM rule as part of the agency's guidance to operators. Eighteen risk factors are outlined, some include illustrative risk values, and a process for combining them into a model is suggested. There are several problems with this (and generally with all risk models of this type) …

The selection of risk factors has no analytical basis. This is not to say that there is no logic or experience behind them. They might be a good place to start investigating real risk factors that should be used in a model. But they are not driven by the data.

They include some factors that we know are less important (less correlated) to risk than other factors that are more strongly correlated with risk, and some that might overlap the information in another factor (like accident history). The test is not whether there is any risk associated with a factor; it's whether these are "good" at reflecting differences in relative risk.

Values for individual risk factors contradict the data. For example, the illustration ranks older pipeline (>25 years old) as a high risk and newer pipeline (<25 years old) as a low risk. The data show that the highest risk is in fact during the first five years after construction or installation, followed by 70-80 years of relatively constant probability of failure. This evaluation reinforces that point, but the point was made in research done (by industry)[75] before IM was implemented. Similarly, historical data show that accident history is a strong predictor of future incidents (think about how auto insurance rates work), but it is not a high/low factor and 10 years is not the break point. The past year is most important, 1-2 years is somewhat important, and the effect fades away after 3 years.[76]

These errors in the model have the effect of *reversing* the risk ranking of systems compared to the actual risk.

Cognitive biases lead to systematic errors—even from subject matter experts and professional statisticians.[77] Research has shown that people misperceive and systematically underestimate certain kinds of risks. While many of the risk factors identified in the rule reflect objective, physical measurements (like pipe wall thickness), others (the importance of previous inspection results, quality of pipe coating, non-standard installation) clearly require judgment.

Experts can become very good at estimating risk, but their judgments generally need to be calibrated first. Calibration may be especially important when dealing with rare, catastrophic risks.[78]

From *The Failure of Risk Management*,
by Gary Hubbard:

When it comes to risks, managers and experts will routinely assess one risk as "very high" and another as "very low" without doing any kind of math. And without deliberate calculations, most people will commit a variety of errors when assessing risks … almost everyone is naturally overconfident in their predictions.

"The degree of this bias is really catastrophic."

Using numerical values to represent categorical variables implies false precision. For example, the kind of product transported must be converted to a risk scale, then converted to a number. But the number doesn't really have the same meaning here as measured values do. In fact, it's not clear how different people would evaluate and quantify the same information.

Using integer values to represent analog values loses resolution in the data at a point in the computation that gets magnified with weighting and combining factors in the model. It seems innocent enough, maybe even necessary if you want a model with consistent scales. But values at the margins (e.g., 3.99 gets rounded down to 3 and 4.01 gets rounded up to 5, where 3 and 5 are the allowable values) get pushed far apart from similar values. Good measurement practice suggests that any rounding should take place at the end of a computation. Otherwise, the results can (and often do) reflect the reverse of the *real* risk ranking.[79]

Scaling the scores for each risk factor distorts the results—the final risk scores—of the model. It implies a linear and proportional change in risk as you go up the scale, such that 5 should be 5 times as risky as 1. If the underlying measure is not linear, the scale should not be. And if different risk factors use different scales, then a 5 in one risk factor doesn't mean the same thing as a 5 in another—which means the scores cannot be added without potentially distorting or even reversing the ranking of risks. And why is 5 the highest value; are there no higher risks?

Weighting the risk factors has no analytical basis. The illustration weights all factors equally. There is clearly no reason why this should reflect the real importance of these factors. An alternative that is commonly used (as in the case of PG&E before the San Bruno explosion) is to weight factors like corrosion or exposure to natural force damage relative to their prevalence in the incident data. But there is no basis for assuming this for any *particular* pipeline, and in fact this method will distort the risks associated with real, local circumstances. Weights from national-level failure data are useful *only if you know nothing else*.

The risk factors are not independent as the model assumes. Age is clearly, logically, correlated with the coating condition, the date cathodic protection was installed, and whether the pipe was hydrostatically tested during construction. Maybe more importantly, some risk factors might be interactive. Known corrosion, for example, might be exacerbated by a disbonded coating or operating stress levels in the pipeline. Manufacturing defects that might otherwise be stable could fail in combination with corrosion or aggressive pressure cycling.

Treating each risk factor as independent potentially double-counts some risk factors where the same underlying condition affects more than one score, and might substantially underestimate the risk from interactive threats.

The model suppresses the importance of risks that might be imminent as they get washed out by many other factors that have nothing to do with the ultimate mode of failure. Systems fail at their weakest link, often (usually) from multiple causes, but not from all causes. The model surveys all the kinds of risk, and in effect, it gives credit for things that are irrelevant. The probability of failure from corrosion or material defect is *not logically reduced* by a damage prevention or public outreach program … or any other factor except one that mitigates that threat. By simply adding things together, the model suppresses the overall risk scores and the range of results so that nothing appears very risky. That might be OK for risk ranking, but it provides a dangerous sense of security for very high risks.

The basic math that glues the model together is inappropriate for ranking risks. Index scoring models use scores from several risk factors, weight the scores, then add them together to reflect an overall risk score. This does not reflect the underlying concept of risk as a function of probability *times* consequences—a multiplication function, not addition. And it does not reflect the underlying concept of failure at the weakest link—zero (safety factor) times zero is still zero. It adds together risk factors that are not using the same scales (apples and oranges), are not independent, and might in fact include interactive threats that are multiplicative.

The math is simple and intuitive, but fundamentally wrong.

We have no way of knowing whether the model actually works—in other words, *whether it actually shows differences in risk*—because the model has not been derived from or validated with actual data. Individual operators will never have enough leak, failure, and accident data to determine whether the model works in predicting failures. And PHMSA doesn't collect the right data to do so either.

Validation is not just nice-to-have. Models used in the insurance industry (and in oil and gas production) are derived from the data and validated with real-world data, because the survival of the company depends on a good understanding of risk.

The quality of the underlying data always limits the quality of the output. There is always the question of how to handle values that are unknown—whether to substitute best estimates, assume a conservative value, or assume the worst case. In some cases, a conservative estimate won't be conservative enough.

Some observations on operators' risk models … from inspectors:

Some companies often don't understand what they are doing with the risk models – might have been better off just using SME judgment.

A lot of high-risk things are grouped with low-risk things so that major issues won't result in high-risk rankings; a lot of "feel good" factors in the risk models.

Companies miss things because it's difficult to link up statisticians with engineering and maintenance groups to develop the models.

Companies don't want to give up relative risk models; they are easier than more quantitative modeling.

Risk models add hardly any value because of such limited data.

The model does not reflect the underlying uncertainty in the information going into it. This is a special problem for a risk model, since the very concept of risk is *a measure of uncertainty*. It's not uncommon to omit a discussion of uncertainty when presenting the results of an analysis, since it is much more difficult for decision makers to deal with a distribution of possibilities than a point estimate. But the uncertainties can be very large, potentially overwhelming much smaller distinctions in the relative scoring of risks, and they can be skewed. A range of $0.6-100 million can become a point estimate of $1 million. Without addressing uncertainty, decision makers can draw the wrong conclusions from the data.

The model has never been "back-tested" to see how well it does in ranking actual risks. The program has 27 years of incident data, including information on many of the variables included in the model that is illustrated in the rule that could be used to check against historical reality. This is not conclusive validation, but models that do not fit historical reality are likely flawed.[80]

It provides the appearance (and comfort) of effective risk management with no empirical evidence that it improves risk decisions at all. There are better approaches …

Evaluating a risk model

Different models have different limitations. Some, like probabilistic risk assessment, are more time-consuming to develop, and they require more data and quantitative analysis. Others are more qualitative, more simple to develop and use, but provide less rigor in the outputs. Many standards and discussions of risk models suggest that the choice should be "fit for purpose."

The British Health and Safety Executive (HSE), for example, compares three general kinds of risk assessments—qualitative (Q), semi-quantitative (SQ), and quantitative risk assessment (QRA).

Figure 6: Approaches to risk assessment

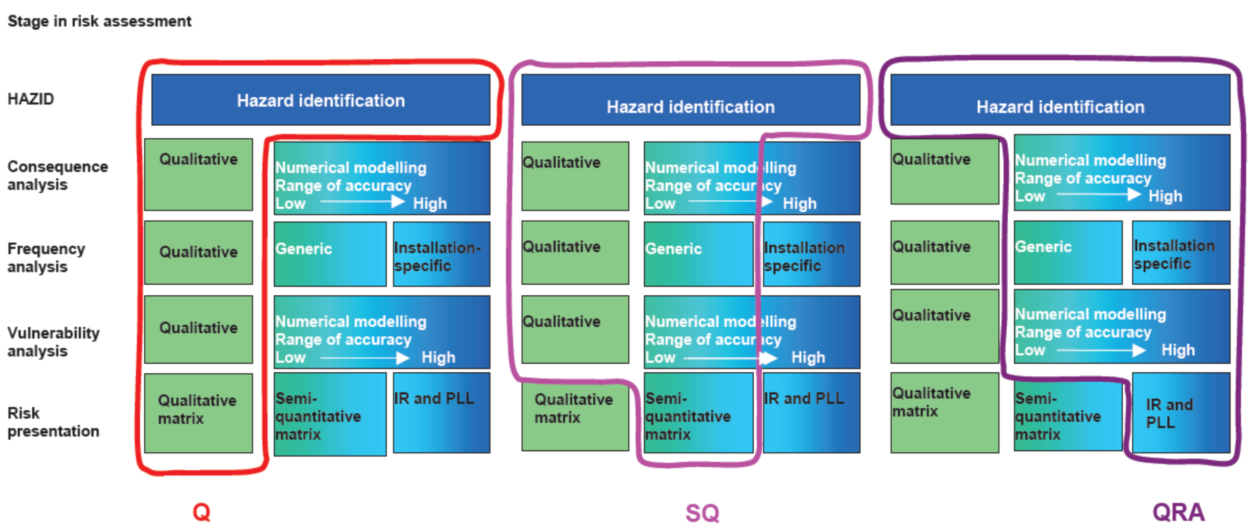


Figure 3 - HSE Guidance on Risk Assessment for Offshore Installations

The HSE guidance suggests that more quantitative risk assessment methods are justified where the hazards are greater. Like many/most other industry and international standards,[81] it endorses the use of index-scoring models as a useful tool for some situations.

To evaluate any risk management method, we have to ask a few basic questions:[82]

- Does it work?
- How do we know it works?
- Would anyone in the organization know if it didn't work?
- If it didn't work, what would be the consequences?

For the IM program models, PHMSA can't answer any of these questions definitively—in the sense of validating or proving the answers with data.

The industry standard acknowledges the need for validation of the model, but the approach is very limited. By comparing results of inspections to determine if the models are correctly characterizing risks, it gets at only whether a particular value or judgment was reasonable, not the overall formulation of the model. And it's not clear to what extent even this is being done.

The problem with an index-scoring model like this is not simply that it works less well than a more quantitative model, but that it might not work at all, or worse—that it might give the wrong results and divert resources from more productive work. It is not surprising that the agency would use a risk modeling approach that is widely used and widely accepted. But it seems that this approach is *widely used because it is widely used*, and because it is comparatively easy. It is in fact widely disconnected from the serious and extensive literature on risk evaluation.

> *Most of the new non-financial methods are not based on any previous theories of risk analysis and there is no real, scientific evidence that they result in measurable reduction in risk or improvement in decisions. Where scientific data does exist, the data show that most methods fail to account for known sources of error in the analysis of risk, or worse yet, add error of their own.*
>
> *Ineffectual methods may even be touted as "best practices" and … are passed from company to company with no early indicators of ill effects until it's too late.*
>
> *- Gary Hubbard, The Failure of Risk Management*

Tversky and Kahneman point out[83] that people rely on a limited number of heuristic principles to reduce the complex task of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics can be quite useful, but sometimes they lead to severe and systemic errors.

Given the very high potential consequences from pipeline failures, it's hard to argue for retaining a simple and seriously flawed risk model.

Probabilistic risk assessment (PRA)—a very data-intensive model—is not the only other alternative. In fact, there are ways to make a relative risk ranking method work. But the risk factors and weights have to be derived from the data—*driven* by the data, not just using data.[84] The math holding the model together has to be guided by the data, clear logic, and sound statistics; the model has to be tested and validated against real, historical data; and an evaluation loop is needed to continuously improve it. This might look like an index-scoring method, but underneath the hood it is fundamentally different.

To be reasonably reliable, this has to be done with a lot of data. No single company has enough failure data or experience to do this; they have a lot of depth and little breadth. And PHMSA generally doesn't collect enough detailed data to identify all the important risk factors—like leak and other failure data, and near-miss data. It has enough to start, and enough to correct some of the information in the appendix to the rule. In fact, this has been demonstrated already—for pipeline safety.[85] But better data on many risk factors could greatly expand and improve the model.

Some general observations on risk modeling

Index-scoring risk models are intuitive, pervasive and widely-accepted, but they all share some serious flaws. These are not simply trade-offs, sacrificing some accuracy for greater ease of use. Index-scoring models have a high potential for *adding risk* into the system and *undermining* sound risk decisions. It's easy to show that this actually happens. But that additional risk is hidden from view, not really obvious from the very quantitative-looking results from the models.

The regulator requires companies to follow industry standards (with limited exceptions). Standards organizations generally have reduced the problem to consensus based on experience. Model users are following a relatively easy, intuitive approach that has the blessing of standards organizations; and their experience is going back into standards. Decision makers get detailed analyses with lots of quantitative outputs, but they have no way of knowing about this underlying risk. Inspectors have readily acknowledged that they are not risk modeling experts, and do not have the expertise to evaluate companies' risk models.

There is no validation of the overall models, no evidence that they work, and no scientific basis for using them. The evidence in fact shows that index-scoring models are inadequate as a basis for targeting IM assessments or any other evaluation of risk. These cannot be fixed by tweaking. The problems are structural and extensive.

Risk analyses should be based upon the best available scientific methodologies, information, data, and weight of the available scientific evidence.[86]

Past research[87] provides a framework for a more useful risk model. Some general principles:

- The purpose is an important starting point. There are certainly differences between how an operator would use a model and how PHMSA would use one. An inspection-targeting model is not simply an aggregation of all the system-level models. Companies should focus on what they can affect (system integrity), and the regulator should focus on what it can affect (company performance).
- Any model should be developed using risk factors that can be demonstrated with data, sorted and weighted based on the data, validated with data, and regularly reviewed and

re-tested against actual results. Experts can suggest risk factors to evaluate, but analysis and statistical measures have to be the arbiter for what's included and what's not; everything else poses a high risk of bias and error. If there isn't enough data to analyze, and a risk factor is believed to be important, the data needs to be collected. This is the general concept of "machine learning."

- Every risk factor needs to take into account the possibility of time degradation, and the possibility of interaction with other risk factors. Combining risk factors is a statistical exercise, not simply adding up the numbers.
- No single company can have a broad enough scope/experience to detect all the patterns and risk factors that might be important. This kind of analysis needs lots more data for multi-factor regression and conditional probabilities.
- The math gluing the model together is critical. Converting categorical and continuous variables into limited scales makes the data easier to handle but it also makes the output *wrong*. Adding risk factors that are not independent and multiplying them all times some mixed consequence factors is intuitively appealing, but also wrong. These mathematical operations can make small differences in risk look big and big differences in risk look small; they can make higher risks appear to be less risky than lower risks; and they hide these differences, so there is no confidence in the risk ranking outputs from the models.
- The range of uncertainty has to be presented to the decision maker. ISO standard 31000 has this point right—decision makers themselves must consider the uncertainties and assumptions in the analysis, and any limitations of the data or modeling.

***Relative* risk models in general have a fundamental weakness**—they don't provide decision makers with any information about how to align spending with the risks they face. They reinforce a steady state approach to integrity management. There may be no sense of urgency when confronted with many higher risks in one company's systems. Risks are simply ranked. This doesn't provide enough information to determine if risks are acceptable or not, given the costs to address them.

Any risk model that is based on historical data alone has another important weakness—it will miss some low-probability, high-consequence risks (those that simply have not yet appeared) and will amplify some of these same kinds of risks (those that have appeared, but might be more akin to the "100-year flood"). It will also tend to underestimate emerging risks. These are special kinds of risk that probably need to be modeled separately based on an expert understanding of systems, failure modes, and risk exposure.

All of this is central to the success of integrity management, as risk models and risk factors underpin so many elements of the program:

- Establishing the baseline assessment schedule (49 CFR 195.452(e), 192.919(c), 192.921(b), and 192.911(c))
- Setting reassessment intervals (195.452(j)(3) and 192.937(a))
- Establishing the frequency of periodic evaluation (195.452(j)(2) and 192.937(b))
- Conducting periodic evaluation (195.452(j)(1) and 192.937(b))
- Identifying the need for, and type of, preventive and mitigative measures (195.452(i)(2) and 192.911(c))
- Measuring program effectiveness (195.452(k))
- Establishing the basis for deviations from the rules (192.913(b)(1))

Data quality

Performance-based regulatory programs tend to be more data-intensive than more traditional, specification-oriented programs. But risk evaluations and models generally are no better than the data going into them.

One of the issues NTSB raised in its investigation of the incident at San Bruno was the lack of original records for the pipe that was used. The operator in that case used incorrect information about the pipe, resulting in a significant underestimate of risk. The broader issue, of course, is the quality—the accuracy, completeness, timeliness, comparability, utility, and relevance—of all the data going into these risk evaluations.

All data systems have error. The goal cannot be perfect data; it's getting data that are sufficiently reliable for the intended purpose. But errors tend to accumulate through many steps in the process—from identifying what we need to know, in the first place, through designing systems, collecting and interpreting data, and ultimately to making decisions based on the findings.[88] It turns out that getting reasonably reliable data is a much more challenging goal than many people realize.

Operators' information about their pipeline systems

Inspectors have observed a wide variety of quality in the records operators have for their systems. They report that a lot of older pipe (before the pipeline safety regulations in 1971) is missing original records, but this varies by company. In the absence of these records, operators might be relying on secondary information—data put on a spreadsheet a long time ago, with unknown reliability, or the opinions of subject matter experts as a surrogate for technical specifications. The IM program appears to have increased operators' awareness of the data they don't have, which may have caused operators to hydro test their systems or find data. Some records were described as "incredibly accurate." But inspectors also noted that records and system knowledge are getting lost when assets are transferred or people retire, and that this problem is naturally degrading more of the data over time.

On May 7, 2012, PHMSA issued an Advisory Bulletin to operators reminding them to preserve and verify records related to MAOP (gas pipelines) and MOP (liquid pipelines). This bulletin also notified operators that PHMSA intends to require gas pipeline operators to submit data regarding mileage of pipelines with (and without) verifiable records in the annual reporting cycle for 2013.

The IM rules require operators to maintain certain records—some for the life of the system. But there is no requirement to transfer records when assets change hands. Operators have a widely-acknowledged obligation to conduct "due diligence" when acquiring an asset.[89] But there is no requirement to transfer records when a pipeline is sold, and both inspectors and industry professionals have observed that records are not always transferred, and due diligence is not always what it should be.

The results from in-line inspection tools introduce several more uncertainties into a risk evaluation. The selection of tools determines to some extent what kinds of defects are likely to be detected. The tool itself will generally miss about 10% of the defects simply based on design limitations. Defects can be mischaracterized, and measurement (length, width, depth) is subject to some amount of error. The interpretation of the results requires specialized expertise to distill what is important. There can be false calls—reported defects that don't actually meet the threshold for reporting. For serious defects, there are calculations of the estimated burst pressure, which have not always matched actual burst pressures from tests or failures. And someone has to decide how to incorporate all this information in a model, and how to present the results for decisions about repair or mitigation.

PHMSA requires operators to consider tool tolerances in their risk assessments, but leaves it up to the operator to decide how.[90] PHMSA has not established performance criteria for ILI tools, and does not intend to.[91] In fact, there is no reason for the agency to do so at this point. The issue is not how much uncertainty is associated with the tool, it's how that uncertainty is treated.

The impact of missing or uncertain data on IM risk assessments varies depending on how the operator chooses to deal with it. Both the API and ASME standards suggest that suspicious or missing data should be flagged so that "appropriate consideration" can be given to it during the analysis process, and they caution against using global assumptions about the condition of a system. Under the standards, ultimately the operator must decide what level of importance will be placed on specific pipeline data. Muhlbauer suggests considering both quality and recency of the data, and distinguishing SME judgments from field verification. He also notes that in the end default values sometimes must be used when there simply isn't any information. The general rule appears to be: *in the absence of good data, use conservative assumptions*, and highlight uncertainties in some way. The San Bruno incident, however, revealed substandard pipe that did not match the as-built drawings and specifications, and that was well below what most people might have estimated using conservative assumptions.

The issue of data quality, in fact, goes far beyond system records. Risk models and risk assessments also incorporate information about the operating environment, which is always changing; results from assessments; and failures. And models don't just use data; they create data. Decision makers use information that has been integrated, analyzed, and interpreted; and each of these steps introduces additional opportunities for error.

Data integration

The IM rule for liquid pipelines requires "*an analysis that integrates all available information* about the integrity of the entire pipeline and the consequences of failure." The rule for gas pipelines requires "an identification of threats to each covered pipeline segment, which must include *data integration and a risk assessment*." An implicit assumption is that operators have data from many sources, have not integrated it, could integrate it, integration would provide useful new insights on risk, and mitigating these risks would further improve safety.

PHMSA provided operators with further guidance on data integration through FAQs and inspection protocols.

The FAQ[92] for gas transmission systems describes data integration as an analytical process considering the synergistic effect of multiple and/or independent facts or data; it highlights an example from the industry standard (ASME B31.8S), to show how coating condition and cathodic protection might be used to evaluate possible corrosion conditions. The FAQ for hazardous liquid pipelines lists some of the more important information that should be considered in an integrated manner:

- Results of previous integrity assessments
- Information related to determining the potential for, and preventing, damage due to excavation, including damage prevention activities, and development or planned development along the pipeline
- Corrosion control information (e.g., test station readings, close interval survey results)
- Information about the pipe design and construction (e.g., seam type, coating type and condition, wall thickness)
- Operating parameters (e.g., maximum operating pressure, pressure cycle history)
- Leak and incident history
- Information about how a failure could affect a high consequence area, such as the location of a water intake

The inspection protocols[93] for liquid pipeline operators ask the inspector to verify that the process for evaluating risk 1) appropriately integrates the various risk factors and other information used to characterize the risk of pipeline segments; 2) uses appropriate variables to adequately characterize the relevant risk factors; 3) has a technically justifiable basis for the analytical structure of any tools, models, or algorithms used to integrate risk information; 4) has logical, structured, and documented processes and guidelines for any subject matter expert evaluations that are used for the integration of risk information; and … 5) integrates any risk model output with any important risk factors not included in the model.

The gas inspection protocols[94] ask the inspector to verify that the individual data elements are brought together and analyzed in their context such that the integrated data can provide confidence with respect to determining the relevance of specific threats and can support an improved analysis of overall risk. Data integration includes a common spatial reference system, and integration of ILI or ECDA results with data on encroachment or foreign line crossings to define locations of potential third party damage.

What inspectors say about data integration:

- *Data integration is a problem. Companies aren't really doing it.*
- *We need to make companies do data integration of pipe attributes. They're supposed to do this, but it's a hodge-podge, not done very well at all.*
- *From failure investigations, we sometimes see some previous kinds of failures that should have been a warning sign.*
- *A lot of companies focused on pigging and digging—don't think they are seriously implementing other requirements like analysis, data integration, prevention and mitigation measures, metrics, and learning about their systems.*
- *Operators are doing the bare minimum, not doing broader risk assessment and data integration; you can see this during inspections and after ruptures, where companies didn't address known problems.*
- *Companies still struggling to get GIS integrated with maintenance systems and other information. We thought we'd get more done on this.*
- *Sometimes operators are missing the forest for the trees—focusing at the anomaly level, not the system level (believe this is driven by dollars). The intent of IM was to get the big picture.*
- *Companies miss things because it's difficult to link up statisticians with engineering and maintenance groups to develop the models.*
- *Performance-based regulation makes oversight more "gray" and enforcement somewhat artful. Some aspects of IM, like data integration requirements, are harder to enforce.*
- *Risk analysis and risk model requirements leave a lot of leeway for operators.*

Challenges with data integration are among the most commonly cited problems with data quality.[95] It's not usually as simple as saying "If I know these two things, then I know something more than either one by itself." Some, probably most, of the data are so extensive that any integration must be automated and modeled in some way to make sense of it. But data are often collected in different formats, for different reasons, and without common identifiers, so effective integration can be near impossible.

Also, just bringing the data together doesn't always help; it can actually be harmful, by burying useful information or averaging it together with much less useful information, resulting in poorer decisions.[96]

Inspectors do not have the specialized expertise or the time to evaluate all of these data quality issues.

Some general observations on data quality

Professional statisticians routinely deal with missing data, uncertainty, imputation, data integration, and data interpretation. This is another domain of specialized expertise not really incorporated into the IM rules and guidance. And many kinds of errors can be amplified and hidden as the data are used in models. Common sense, familiarity with the data, and other technical skills can provide a false sense of confidence in dealing with these kinds of data quality issues.

Perfection in data is not a reasonable or useful aim. Skilled handling of the data is. And that is also the path toward continuous improvement. DOT's Information Dissemination Quality Guidelines (2002) provide some useful material to help evaluate and improve data quality.

Metrics

Metrics (or indicators) can be a powerful engine for improvement, and a way to gauge whether improvements are actually occurring. Metrics can be used to monitor and help understand:

- the condition of the system
- changes in risk exposure
- processes and activities – to help track implementation of the program
- immediate or intermediate outcomes – the direct results of activities
- safety outcomes – to see if the expected results are occurring, as a benchmark for evaluating program effectiveness
- deviations in the normal functioning of a system, that might indicate increasing risk
- patterns – to detect emerging trends, or to help identify important risk factors

NTSB, in its report of investigation on the San Bruno pipeline explosion, recommended that the pipeline safety program expand the use of meaningful metrics—to develop and implement standards requiring operators to regularly assess the effectiveness of their programs using clear and meaningful metrics, to identify and correct deficiencies, and to make those metrics available in a centralized data base.

The measures and data we have now

The IM rules require operators to measure whether the program is effective in assessing and evaluating the integrity of each covered pipeline segment and in protecting HCAs. From this starting point, the liquid and gas rules diverge.

The liquid pipeline rule includes some examples of performance measures an operator could use to meet this requirement:

- Total volume from unintended releases (with a goal to reduce by some percentage)
- Total number of unintended releases of 5 gallons or more
- Percentage of integrity management activities completed each year
- Effectiveness of the operator's community outreach activities
- A narrative description of pipeline system integrity, including a summary of performance improvements
- A performance measure based on internal audits of the operator's pipeline system
- A performance measure based on external audits of the operator's pipeline system

- A performance measure based on operational events (e.g., relief occurrences, unplanned valve closure, SCADA outages, etc.) that have the potential to adversely affect pipeline safety
- A performance measure to demonstrate that the operator's IM program reduces risk over time with a focus on high risk items
- A performance measure to demonstrate that the operator's IM program for pipeline stations and terminals reduces risk over time with a focus on high risk items

The gas rule requires four specific, overall performance measures:

- Number of miles inspected vs. program requirements
- Number of immediate repairs completed as a result of the IM inspection program
- Number of scheduled repairs completed as a result of the IM inspection program
- Number of leaks, failures, and incidents—classified by cause

… and another 32 threat-specific metrics (across nine threat categories) outlined in ASME standard B31.8S. The standard also provides an additional nine leading and lagging indicators (from activities through direct integrity measures), and another 14 overall performance measures that operators should use.

None of these, individually or collectively, actually measure program effectiveness. The distinction is important here because *evaluating the program*, in a process aimed at continuous improvement, is one of the key elements of a safety management system. Evaluation is a process of clarifying program logic and criteria for effectiveness, testing the assumptions, and developing objective evidence to help understand the results. The IM program doesn't take this beyond collecting and reporting summary level indicators. There is no real feedback loop to assess and adjust—which is far more important than counting numbers.[97]

Without clear guidance on how to measure effectiveness, the program is likely to have somewhere between 3,000 different methods (which it could not evaluate effectively) or none (which is more likely without a requirement beyond collecting data for some indicators).[98]

Annual reports to PHMSA require a range of performance measures. All are to be broken out by commodity group (crude oil, refined non-HVL products, HVL, CO2, ethanol, natural gas, synthetic gas, hydrogen gas, propane gas, other gas):

- Pipeline miles in [or that could affect] HCAs (onshore, offshore)
- By interstate or intrastate (separate reports for each) –
    - Mileage inspected – broken out by ILI tool type, direct assessment, or other inspection technique

- Actions taken based on ILI inspections
    - Anomalies excavated
    - Anomalies repaired based on operator's criteria
    - Conditions repaired within an HCA (immediate, 60-day, 180-day, one year, monitored, and other scheduled conditions)
- Mileage inspected by pressure testing
- Actions taken based on pressure testing
    - Ruptures and leaks repaired (total—both within and outside an HCA)
    - Ruptures repaired within an HCA
    - Leaks repaired within an HCA
- Assessment miles completed during the calendar year (in HCAs or segments that could affect HCAs only)
    - Baseline assessments
    - Reassessments
    - Total

- Incidents, failures, and leaks eliminated/repaired – by 13 causes – for gas pipelines only (reported separately in accident reports for liquid pipelines)
    - Transmission incidents in HCA segments
    - Transmission leaks (onshore, offshore) – HCA, non-HCA
    - Transmission failures in HCA segments
    - Gathering leaks (onshore, offshore) – Type A, B (for onshore only)

None of these metrics provide enough detail to be useful. Summary data actually adds to the reporting burden and the potential for error, as companies must process the data they have in the form they have it, to give it to the regulator in another form. But the biggest problem is that it makes the data near useless for any analysis.

There is no way to connect failures to defects, defects to pipeline attributes or risk exposure, or to identify any of the risk factors that might be important in evaluating and reducing risk. We can't say much about the condition of the overall system, or the relative risks across different systems. We don't know anything about deterioration rates, and we can't track conditions over time. We don't know how well assessments are working, and we don't have any leading indicators of emerging risk. We don't know where repairs are being made.

The data are not used—and in fact cannot be used—by PHMSA for much beyond superficial monitoring. The data provided little help or insight in answering the questions in this evaluation about program effectiveness, or why we're seeing what we're seeing. This is a problem not only for program management, but also for the companies' risk evaluations. Identifying and quantifying risk factors requires far more data than any one company has. The lack of data/analysis at a national level leaves all companies in the dark on estimating risk—accurately and consistently—for nearly every risk factor PHMSA requires companies to consider.

Measurement concepts

Good measures should start with the concepts we want to measure—*What do we need to know?*[99] Here are some things both the regulator and operators probably need to know to manage the program effectively—a provisional list to help frame the concepts:

- ***What really causes systems to fail?*** What are the important factors, conditions, circumstances, and events? How do they interact in a way that leads to failure, or perhaps amplifies risk?

  PHMSA needs to know this to target research and options for intervention (including standards or regulations), to develop risk factors—with a demonstrated relationship with real risk—that can be used in risk models/evaluations, and to focus program inspections.

  > ***There are not enough trees in the rain forests*** *to carry all the procedures necessary to guarantee safe operations.*
  >
  > *Safety depends crucially on a clear understanding of the interactions between many different underlying processes.*
  >
  > - James Reason in Managing the Risks of Organizational Accidents (p.181)

  Some things the program might measure/track: All the conditions and circumstances for every failure, in addition to the "causal chain" or interactions that ultimately led to the failure, and going back as far as possible in the chain; conditional probabilities for various risk factors and multi-factor regression to tease out relationships and importance.

- ***Where do these conditions exist now, and how is this risk distributed?*** Which systems, parts of systems, regions, companies, etc. are at higher risk? Which of these conditions or events are inherent risks (associated with the physical system) vs. performance risks (associated with the management, operation and maintenance by a company)?

  The program needs to know this to understand where risk is concentrated, to assess the costs and benefits of various alternative interventions (including regulations), to target companies or systems for inspection, and to track changes in risk as assets change hands.

  Some things PHMSA might measure/track: Condition of the system (part of relative risk) – by segment, including information about all the latent conditions, with enough detail to evaluate against all the risk factors that have been identified; maintenance findings; the safety and environmental outcomes (failures) that are actually occurring.

- ***How are these conditions changing over time?***

  The program needs to know this to forecast resource requirements and allocation of resources, to help understand trends in safety/environmental outcomes, and to target future inspections.

  Some things PHMSA might measure/track: Deterioration rates (average and distribution) for time-dependent threats – to help set re-inspection intervals; this can be for physical things like corrosion, but also for organizational things like training and even PHMSA inspections of the overall quality of systems. Trends in the risk factors, including forecasts wherever possible. Repairs made – with enough information about them to evaluate against all the risk factors that have been identified, and to permit comparison with later failure data.

- ***What kinds of controls or barriers[100] are effective in reducing or mitigating risk?*** What physical controls (e.g., relief valves, automatic shutoff valves) or operational controls (preventive and mitigative measures) work? How well do they work, and how do other conditions affect this?

  The program needs to know this to compare intervention strategies (including regulatory alternatives), to share lessons learned with operators, to evaluate the effectiveness of regulations, and to help evaluate operators' programs.

  Some things PHMSA might measure/track: Preventive and mitigative measures put in place – and where – so their effectiveness can be tested and evaluated on a large scale, and so that the regulator can begin to assess operator performance as input into the inspection process; the relative strength of safety culture in each regulated company; valve types and spacing.

- ***What's the status or condition of these barriers in relation to the threats (system by system, segment by segment)?***

  The programs needs to know this to help evaluate the effectiveness of barriers, to help target and focus inspections of companies and systems based on risk, to monitor overall program risk, and to help understand changes in risk concentrations. Companies need to know this to manage their risks effectively.

  Some things PHMSA might measure/track: Status of barriers or controls – availability of relief valves, etc.—by system and segment; numbers of alarms (correct and false); levels of training and qualifications, and measures of change in these over time; deviations from procedures (which could be either good or bad); reports of inadequate tools or equipment or unworkable procedures.

- ***How can the program detect emerging risks or latent, catastrophic risks?***

  Program managers need to know this to develop barriers before a major accident. Very low probability high consequence (VLPHC) accidents tend to be different in nature, so lots of routine incidents are not necessarily a good predictor. Need to estimate frequency of these (for regulatory and policy evaluations), and isolate the risk factors from accident investigations and analysis.

  Some things PHMSA might measure/track: Near-misses or partial failures – e.g., component failures, alarms, overpressure events, etc. – these are early warning signs; every failure or indication of trouble – the kind that often precede major accidents (e.g., leaks, or any unintentional release); high levels of workload that might reveal increasing pressures to cut corners; deferred maintenance; changes made in systems, maintenance, operations, or management.

- ***Do the programs work in reducing risk?*** The IM rules require operators to assess the effectiveness of their programs, but the indicators that are suggested in the rules don't reflect the analytical requirements to evaluate program effectiveness. There is no regular, ongoing evaluation of the national program, and there was no evaluation built into the program design or implementation.

  The program needs to know this to adjust/adapt its programs, to tailor them based on local conditions, to keep the costs as low as possible for the greatest return on investment, to eliminate ineffective requirements, and to target interventions.

  Some things PHMSA might measure/track: The extent of program implementation (assessments done, conditions found, the results of evaluations, repairs made or not—all at the segment level); compliance history from past inspections; and the ability to back-track to these for every failure that occurs.

Failure data provide a core litmus test for judging risk. Small failures (leaks, activation of relief valves, safety-related conditions, etc.) signal weaknesses that might indicate a risk of a larger failure. These are in the domain of the operator, looking at the condition of its own system. But the *rate* of failure is critical for identifying and weighting any risk factor in a model. No single operator has enough failure data or experience to determine the importance of these risk factors in a reliable, quantitative way. There are several special problems with the failure data the program collects:

- *The accident/incident reports don't provide enough information about causes and circumstances to identify important risk factors*. Most accidents result from a combination of many latent conditions, circumstances, and triggering causes. In the incident in San Bruno, the operator reported the cause as *material/weld failure-original manufacturing-related (not girth welds or other welds formed in the field)*. In contrast, the NTSB report identifies multiple kinds of material/weld failure (including welds in the field), corrosion, fatigue, operations/maintenance error, control room communications, and many other contributing factors to the accident. By narrowing the data to a single cause, we end up very little understanding of what actually happened, and almost no ability to find patterns in the data through statistical analysis.

- *Cause codes are near-useless*. Human error is a consequence, not a cause.[101] As one of seven overall cause codes, it reflects a cognitive bias (fundamental attribution error), and provides no help in program improvement. Material failure is a consequence, not a cause. So is excavation damage. Sub-cause codes are an attempt to elaborate on these, but they almost all get at what failed or how it failed, not *why* it failed.

- *Outcome indicators don't necessarily reflect real changes in risk.* They provide an ultimate test over a long period of time, but they do not provide a reliable indicator of a systems' intrinsic safety.[102] Small numbers, especially can mask the kinds of risk we are most interested in—low-probability, high-consequence risk.

- *The program doesn't collect data on precursor failures* to estimate the probability of LPHC events, or to compare the risks of different systems effectively, or to detect emerging risks.

- *The reporting thresholds themselves* make the data unusable for certain kinds of analysis using conditional probabilities.[103]

Pipeline attribute data (things like diameter, material, year of construction, soil type) provide descriptive information about the system. These attributes don't cause accidents, but they are commonly associated with different level of risk, so they might be important risk factors.

- *The attribute data in annual reports are all summary-level*—processed data where the useful detail has been removed. We can't analyze the interactions of various risk factors, because the summary data don't (and can't) tell us all the possible combinations we would want to consider.

- *Attribute data in annual reports don't match the data reported through the National Pipeline Mapping System,* and the differences are not easily reconciled. This creates a real impediment to data integration.

- *Data in the annual reports cannot be integrated with incident data* because the data are reported in different forms, with different data elements, and at different levels of summarization.

Pipeline condition data potentially give us a picture of the integrity of the system, the nature and extent of deterioration, and the ability to monitor changes in condition over time.

- *We have virtually no information about latent conditions* (defects, poor design, clumsy automation, etc.) that increase system risk, or about the condition and performance of barriers (pressure relief valves, alarms, etc.) that are supposed to help reduce risk.

- *Condition data are all provided at a summary level*. For example, the program collects data on the number of pipeline miles inspected and the number of repair conditions detected each year, but the data are not useful for comparing systems or looking at parts of system based on certain attribute information.

Pipeline operations and maintenance data could give us a picture of what is being done to mitigate threats, as a window into the effectiveness of different preventive and mitigative actions. This might also help evaluate the relative performance risk across companies as a way to target the highest risks in IM inspections.

- *Very little O&M data are reported* to PHMSA, and what is reported (for example, numbers of repairs made) is at such a high, summary level that it is not useful at all for analyzing or targeting risk.

The program needs to measure/monitor both vulnerability and resistance to failure. Systems are more or less vulnerable because of the inherent risks in design, latent defects, deterioration over time, changes that can increase complexity and reduce visibility at the same time, lapses in maintenance, new defects added through maintenance, poorly-designed automation, and the human element. Systems are more or less resistant to failure because of physical separation from threats (depth of cover, right of way, etc.), backup systems, procedures, fail-safe mechanisms, maintenance, *and the human element*.

The program has virtually no data on any of these. No single company has the ability to measure the risks and effectiveness of countermeasures in any meaningful way, and there is no national data set that might be used to do that.

The program needs to monitor all the little failures that precede or indicate bigger failures. This is a key feature of almost every safety management system or performance-based regulatory approach. Exxon calls these "free lessons." But to the extent that it's being done, mostly it's done small—in the domain of the individual operator. So these little failures serve only as "dummy lights" on the dashboard. They warn of a problem, but there is no gauge to indicate where on the risk continuum the warning falls. This is the same problem the program has with monitoring vulnerability and resistance. No single company has the data to effectively measure the risks, and nobody is collecting it at a national level. For this to work, the data need to be disseminated widely.[104]

This is essentially what NTSB is recommending—a national data base. The quality of every company's risk evaluation depends on it; and so does the effectiveness of the national pipeline safety program. There are certainly challenges in trying to collect these kinds of data. There are conceptual challenges, technical (IT and data base) challenges, and administrative challenges (requirements of the Paperwork Reduction Act). There are challenges in finding *meaningful* data within a much larger set of data. But the need for grounding the program in good quality data is compelling, and the current state falls far short of providing the data the program needs to manage effectively. It has to start with a broad re-thinking of information needs. That should be followed by exploring alternatives (who might collect and analyze the data, how it might be assembled, etc.). It might not have to be a government data base.

Performance goals

Goals can focus effort and drive changes in performance. It's commonly said that "what gets measured gets done."

Goals cannot be used as a substitute for the regulatory process. The Administrative Procedures Act requires a systematic process for setting standards or requirements companies must meet, with opportunity to comment and provide evidence, and generally a need to justify requirements with an economic analysis of costs and benefits.

Of course, there is a lot of space in between. An agency might:

- *establish a requirement that all companies set performance goals* in certain areas, develop logical strategies for meeting the goals, evaluate external factors that could affect performance, monitor progress, report the results, and explain what actually happened. This is the kind of accountability that the Office of Management and Budget has recommended for agencies themselves in Circular A-11;

- *use goals to help correct an unsafe condition*. For pipeline safety, goals are sometimes used in Corrective Action Orders (CAOs) following a failure or some other detection of an unsafe condition. CAOs can include conditions for resuming normal operation, and intermediate goals can be a way to track progress and expedite a return to normal, safe operations;

- *set goals in a regulation to expedite implementation*. The IM rules established timeframes for completing baseline inspections, including a goal of completing 50% within three years, as a way of expediting implementation;

- *encourage companies to set and monitor goals to help establish and maintain a strong safety culture;* and/or

- *use the fact of voluntary goal-setting and performance against goals to help target inspections and/or report operator performance to the public*.

The IM rules include a requirement for measures, and the attached guidance clearly shows an expectation that these would be used to set goals. The examples, though, were all expressed in terms of changes, not meeting any absolute standard. And in the inspection/oversight process, there is *no accountability for meeting or not meeting any goals*.

Criteria for performance measures: The history of developing and using performance goals for Federal government programs provides some useful considerations if PHMSA wants to use goals to help influence operators' performance. From that experience, agencies have learned that "good" performance measures:

- *communicate value – they* provide a direct indicator of an outcome that people care about;
- *are easy to understand* and relate to; they stand alone as a measure of success – not dependent on complex interpretation; they are easy (conceptually) to construct, and they match a common sense understanding of what's happening in the real world;
- *are measurable with reliable and reasonably available data*; these might vary across different types and sizes of operator;
- *are credible externally,* not easily "gamed;"
- *are directional and sensitive to real changes* – allowing an assessment of whether things are really getting better;
- *are independent of the strategies used to achieve them*, permitting the widest range of options for action to achieve results;
- *are analyzable,* subject to progressive disaggregation to better understand what's happening, help diagnose problems, and identify areas of focus. This too might vary with the type and size of system;
- *are normalized* for changes in exposure if possible, or shown with separate exposure indicators, where applicable;

- *set the right incentives* for program managers (they don't distort priorities); and
- *are reasonably ambitious*, to help motivate new ways of thinking.

James Reason suggests that the key is in appreciating what is manageable and what is not.[105] He suggests managers should focus on regularly measuring and improving processes—design, hardware, training, procedures, maintenance, planning, budgeting, communication, goal conflicts, etc.—that are known to be implicated in organizational accidents.

There is a strong argument for both outcome and process measures. Outcome measures keep the focus on results that matter. With analysis and a good diagnostic approach, they provide a window into what's happening, and how risk might be changing. The big impediment is the problem of accountability.

Accountability: The federal experience also suggests that where goals are *outcome*-oriented (i.e., subject to *influence*, but beyond an organization's *control*), *accountability* cannot be in the strict sense of pass/fail. In that case, accountability should be based on an organization's:

- focus on important outcomes, and ability to develop logical priorities to affect these outcomes based on data and other information;
- monitoring the effects of actions taken and adjusting during the year and continuously improving processes and developing creative solutions;
- understanding of the external factors that might also influence results; and
- explanation of the results at the end of the year, including evidence of how the priorities and actions affected the outcomes.

But this is really just doing integrity management.

Inspection/oversight

Safety regulatory agencies generally achieve results by setting and enforcing standards, which requires a program to inspect against the standards. These are core activities, often accompanied by many other activities (like research, analysis, outreach, training, grants, and information dissemination) to help achieve compliance and good safety practices.

Inspection programs generally include some means of training and qualification of inspectors; a method for scheduling or targeting inspections—either on some regular, recurring basis (like every 3 years) or targeted based on some formula; some pre-inspection review of data/information to help focus the inspection; inspection checklists, protocols, and other guidance for conducting an inspection; some reporting mechanisms; and enforcement guidance for dealing with deficiencies.

NTSB, in its report of investigation on the pipeline incident in San Bruno, questioned whether the IM inspection protocols were adequate for:

- ensuring the completeness and accuracy of pipeline operators' integrity management program data, and
- ensuring the incorporation of an operator's leak, failure, and incident data in evaluation of the operator's risk model.

This evaluation has discussed many of the data quality issues affecting operators' IM programs and risk models. But these are not simple protocol issues, and the answer is not just a better protocol. Identifying data quality issues and judging appropriate use of the data requires time, specialized expertise, and a forensic approach to evaluation.

How IM has changed the inspection process

Performance-based regulations bring a number of special challenges for inspection and enforcement. These generally change the process to more of an auditing function, requiring a somewhat different (and overlapping) set of skills. And they can greatly increase the knowledge requirements for inspectors.

Inspectors generally acknowledge that IM requirements are much more difficult to inspect against. They noted that IM requires higher levels of qualification, longer training and experience, and (for many requirements) a high degree of specialization. One inspector commented: *"For a lot of IM requirements, you need a specialist. Companies have specialists; we don't."*

This concern about specialization extends to risk modeling ("*We're not risk modeling experts*"), risk analysis, program evaluation ("*sometimes it doesn't look right, but we don't have an alternative*"), data quality, and metrics. One inspector also noted that they don't fully understand the financial side of the picture—the incentive structure that drives behavior in companies.

Inspection protocols

In implementing the IM rules, the agency developed detailed inspection protocols, and a risk model[106] to prioritize gas transmission operators for IM inspection. The protocols, in particular, were designed to guide inspectors through an investigative approach in assessing compliance with the IM regulations.[107] The protocols fill 121 pages, cover 8 issue areas, 46 separate protocols, 304 characteristics to look for, and reference a large body of technical reports to help inspectors determine the adequacy of each IM program. But clearly inspectors believe they need still more guidance to audit operators' programs effectively.

Inspectors reported that it's hard to determine if an operator has done an adequate job with prevention and mitigation measures, hard to evaluate risk-based reassessment intervals, and generally hard to apply the guidance to smaller operators. They don't necessarily want to be in a position to tell operators how to do things, but need to be able to evaluate what an operator is doing.

In fact, there are many deeper mismatches here ...

- The inspection protocols require inspectors to evaluate a company's risk models, but inspectors readily acknowledge that they are not risk modeling experts.
- The protocols require inspectors to assess data quality and the statistical treatment of missing data and uncertainty in the risk evaluations, but these are statistical issues in another domain of expertise; inspectors are not statisticians.
- The protocols require inspectors to judge the effectiveness of a company's processes to manage change in the organization, but they have no particular expertise in organizational behavior.
- The protocols require inspectors to evaluate whether a company's performance measures adequately measure the effectiveness of their IM program, but they have no particular expertise in performance measurement or program evaluation.

As James Reason points out, "Front-line regulators are generally technical specialists, but major accidents arise from the unforeseen—and often unforeseeable—interaction of human and organizational factors. In this new climate … regulators are required to look out for deviations of a different kind, with requirements expressed in far more general terms, that can vary widely from organization to organization."

Inspection Protocol 5.02, for example, provides guidance in evaluating the integration of information in a risk analysis (italics and underlining added):

> *"An effective operator program would be expected to have the following characteristics: 1) inclusion of the <u>appropriate</u> variables to <u>adequately</u> determine the <u>relevant</u> risk ranking of a pipeline segment, 2) a <u>technically justifiable</u> basis for the analytical structure …, 3) <u>logical</u>, <u>structured</u>, and documented processes and guidelines …, 4) <u>justification</u> for …, 5) a process that <u>emphasizes</u> potential risk …, and 6) a method that <u>integrates</u> the risk model output with any <u>important</u> factors … to provide a <u>more complete</u> evaluation of the risk."*

Each operator has its own program with different processes and standards; there are hundreds of operators; and the inspector's judgment must compete against the professional judgment of the operators, while inspectors generally acknowledge they can never have the level of specialized expertise operators have. This is a difficult position for an inspector.

How integrated inspections will change the process

The Office of Pipeline Safety has been working for several years toward an *integrated inspection* regime to pull together all the major types of pipeline inspections into a more comprehensive, risk-based (or risk-informed) approach. The idea would be to move away from periodic inspections (every 3-4 years) toward risk-based targeting of operators—using data from incidents, inspections, pipeline characteristics, and annual reports from the operators themselves. And as inspectors conduct these new inspections, they would look at the nature of the risk for each operator, beginning with a broad assessment and drilling down as they found issues during the inspection.

The general logic for this initiative, as outlined by program managers:

- The program doesn't have the resources to continue doing business as it has been.
- Over the last decade, PHMSA has been adding new inspections to address new issues, without stepping back and rethinking the whole process.
- The pipeline safety program overall is reaching diminishing returns in safety, and needs to focus more resources where the risk is to achieve program goals.

- Operators with good safety programs shouldn't need as much scrutiny as those with higher risk.
- The current practice of separate, disconnected inspections doesn't provide a comprehensive picture of an operator's risk.
- Team inspections can be powerful as they bring together different skillsets to help evaluate many disparate elements of an operator's program.

Field inspectors and region directors expressed some serious concerns about the transition to integrated inspections. Fewer than 20% of the people interviewed described this in positive terms. Some of the concerns people expressed:

- To get good safety performance from operators, they need to know we're coming back on a regular basis.
- The program doesn't have the data or analytical basis for targeting risk-based inspections.
- The big *program* check (from IM) will disappear as inspectors move from looking at operators to looking at *systems*; programs can include several systems.
- The average experience level of inspectors (looking at IM issues) will be lower, as all inspectors will do all inspections.
- The scope of a comprehensive inspection is too large—it will keep you from really drilling down and looking at IM results.
- We're trying to cover too much with mega-systems.
- There isn't enough good source data to efficiently determine what questions to ask.

One of the biggest issues here is the relative advantage of targeting inspections vs. regularly scheduled inspections. A risk-based program has some intuitive appeal. Everyone would probably agree that resources generally should be focused where the risk is. But there are several challenges with this approach.

> *There is no evidence that a risk-based approach will get better results*. There is a fundamental divide on whether an enforcement program is better designed with regular inspection intervals or irregular, risk-based targeting; and how companies will behave under either approach. This has serious implications for safety risk. But there has been no evaluation of the effectiveness of the current (regular) approach before moving to a new one, nor is there any test or evaluation built into the new program.

There is no evidence that a systems approach will get better results than unit-focused or operator-focused inspections.

A substantial fraction of the organization doesn't understand or support the changes. One of the key factors in the success of any program is a broad base of organizational support. Right now, the breadth and depth of concern in the field alone might be enough to threaten an effective transition in the inspection program. Management of change is an important concept in PBR; the program has not applied it in this broad transition that is planned in the inspection program.

Our risk model and data are inadequate for risk-based decision making. The Risk Ranking Inspection Model (RRIM)—the basis for targeting risk in integrated inspections—shares most of the same structural flaws as the index-scoring models companies are using. And much of the data the program collects are not useful for identifying and weighting important risk factors in a model.

There is probably a useful middle ground here—better *using* risk data as input into targeting or focusing inspections, but *not limiting* the efforts to those areas that appear to be the highest risk. It is also important to evaluate/compare the relative value of these inputs to validate the models and make improvements over time.

Evaluating States who inspect for compliance

States inspect the large majority of pipeline miles in the U.S., including a substantial fraction of the transmission mileage. States have about 80% of the pipeline inspection workforce, and the systems they regulate are involved in about 80% of the deaths and major injuries. Now that integrity management has been applied to distribution systems, States have a greatly increased role in overseeing IM programs.

This presents a complicating factor, of course, as PHMSA is one step removed from the process. Another complication: for intrastate gas transmission systems (such as the pipeline in San Bruno), states are not enforcing PHMSA's regulations, they are enforcing their own[108]—which must be at least as stringent as the federal requirements. As a practical matter, they use the same inspection protocols for their IM inspections, and the same training, and often IM inspections have included both federal and State inspectors each looking at the systems within their respective jurisdictions.

Inspections *vs.* root cause investigations

Inspections are largely about compliance; one of the principal purposes of an inspection is to determine the effectiveness of an operator's program. The IM rules also require operators to periodically evaluate the effectiveness of their programs, and one of the ways PHMSA suggests doing that is with root cause analysis of failures and near misses. The FAQ[109] asks: *Are these occurrences being critically examined and are the lessons learned being implemented?*

It seems logical to ask this same question at the national level. Why limit learning to the company that experienced the failure? What about patterns of failures that are not apparent in a single accident—who would notice these if the analysis is limited to each company's own experience? What do the root causes of failures suggest about the effectiveness of the existing rules and standards? What do they suggest about an operator's performance relative to other operators? What do they tell us about the tool capabilities or limitations, data interpretation or presentation, or rates of deterioration for time-dependent threats?

PHMSA does a limited number (about 35, on average) of root cause investigations each year. For all other accidents and incidents, the agency has either:

- No data (for failures below the reporting thresholds),
- Very limited data (for safety related conditions that are not fixed within 5 days), or
- Limited data, with every accident or incident narrowed down to a single cause.

This is not enough to distill important risk factors for pipeline safety, and often it is not enough to support regulatory action even for known problems. It is not enough to discriminate between operators who are managing their systems effectively and those who are not. And it is not enough to detect emerging trends or leading indicators of safety problems.

One of the ideas suggested by several inspectors is to re-balance the emphasis of pipeline safety from inspections to more investigations of accidents and failures. Clearly not all agree. But there is a strong connection between this idea and the general approach of performance-based regulation. To oversee risk management programs, PHMSA needs to understand risk well. It needs to understand better how organizational factors interact with other risk factors to cause accidents. It needs to watch for emerging trends. It needs better data on what really constitutes a risk factor for pipeline safety. In fact, many of the problems with the risk models, data quality, and metrics cannot be solved without a better understanding of how failures are occurring.

Some comments from inspectors:

We should be doing more failure investigations and fewer inspections; all of this should be counted as "inspection time."

PHMSA doesn't have any leverage before an accident happens, but has lots after an accident – we should consider reversing the relative investments in inspection vs. investigations.

And the program needs to be able to evaluate the effectiveness of its standards and inspections. When failures occur, where was the gap? Right now, in most cases, we don't know.

The program has 135 inspector positions today, including five designated accident investigators, distributed across the five regions. Many other inspectors get involved in accident investigations, and collectively they investigate an average of about 35 accidents per year to determine root cause. The agency also has asked for more investigators in its FY 2013 and 2014 budget requests (the latter of which is pending Congressional action on appropriations). These might be a good start. But an effective investigations program needs several other key elements, which remain to be fully developed and implemented:

- program *guidance* for conducting good root cause investigations;
- a *good conceptual model* of failures and *a data base* that is designed to reflect all the causes and circumstances associated with an accident, in a way that helps reveal interactions in various risk factors and how barriers or controls might help;
- a process for *engineering review* of each investigation report to distill safety issues and good practices that might be more broadly relevant to other companies;
- a process for *statistical review* of investigation data to help identify patterns and trends that might be hidden in individual accident details, and to analyze risk factors that can be used in models.

Some general observations on the inspection process

Inspections generally go beyond evaluating for compliance with the rules. They typically include also discussion of safety issues with company personnel, educating them on the requirements, learning about company practices and lessons learned, sharing practices from other operators, encouraging practices that go beyond regulations, and collecting data/information to build the knowledge base. The inspector's job is complex.

Inspection oversight in a performance-based regulatory environment is inherently more complex than for specification-oriented rules. By design, PBR leads to a wide array of tailored processes and systems to achieve a common result. Companies can build specialized expertise around their own processes and systems. Inspectors are faced with the need to understand all of them at some level, and the need to go much deeper into the data. They clearly feel comfortable doing this in some areas more than in others. The program has concentrated many of its more seasoned inspectors on IM oversight to help address these challenges. But it might be near the limit of effectiveness with existing processes, skills, and resources.

Integrated inspections (II) could provide a more forensic approach to inspection oversight. The general idea is to begin at a fairly high level, look for indications of problems, then dive deeper to explore as needed. And the use of teams broadens the pool of expertise. This approach seems more consistent with a PBR environment. But it brings a range of implementation challenges of its own. It is also coupled with another change—moving from periodic inspections to risk-based targeting—that presents some more serious difficulties. If these two changes could be de-coupled, II might be easier to navigate to a successful program. Risk-based targeting needs considerably more evaluation of the program logic, program data, risk modeling, and inspector concerns. It should also be based on reliable baseline data measuring the underlying rate of compliance (all companies, apart from any targeting) as a point of departure for targeting.

Many in the field see a large gap between expectations for the inspection process (to get it done right) and the resources allocated to it. In 2011, about 11% of the agency's inspector resources were spent on IM inspections, and that was up from previous years.[110] There also may be a gap between where the program has inspection resources and where the highest priorities are for inspection. In interviews conducted for this evaluation, several field staff suggested redistributing or reorganizing pipeline safety inspection resources based on where the risk is. But before reallocating resources based on risk, the agency's risk model needs to be re-worked and validated to address the issues above. And program managers might consider other alternatives (like third party audits) for auditing pipeline operators' programs.

The balance between inspections and investigations might seem like a simple, discretionary management choice—a matter of preference. Probably neither function has enough resources to do it as well as people would like. But going back to fundamentals, it seems the first and most basic element of a safety regulatory program is intelligence. The program needs a sound understanding of how and why things actually fail as the foundation for rules and program priorities—and these in turn provide the foundation for compliance inspections, oversight, R&D, and every other program activity or interest. The pipeline safety program does not have sufficient data for statistical analysis to identify risk concentrations and detect emerging trends on how and why pipeline systems fail.

Enforcement

The primary purposes of enforcement are *correction* and *deterrence*—to correct unsafe conditions, and to raise the cost of non-compliance so that companies comply with the rules and follow safe practices. Ultimately, this is about influencing people in the regulated community—influencing their values, actions, and behaviors in such a way that they comply with the rules and demonstrate a commitment to safety over competing demands.[111] One of these competing demands, of course, is profit or financial gain.

The pipeline safety program has many tools in the toolbag for achieving compliance and influencing companies' safety programs. These include:

- *Clear, simple, and cost-effective rules* – the starting point is a set of standards that are relatively easy to comply with
- *Education* – filling any knowledge gaps, including sharing good practices, findings from accident investigations, issuing safety bulletins about technical issues
- *Persuasion* – making technical arguments, giving operators room to comply before more serious enforcement tools are considered
- *Civil penalties* – imposing a monetary cost for non-compliance, beginning with a Notice of Probable Violation, through a process managed administratively by the Enforcement Division in OPS and the Office of Chief Counsel
- *Criminal penalties* – imposing more severe sanctions for willful violations, pursued through the Department of Justice
- *Notices of amendment* – requiring changes to a company's program through a formal process
- *Corrective action orders* – imposing conditions for operation, often used after an accident
- *Compliance orders* – requiring positive compliance by certain dates
- *Safety orders* – imposing conditions for operation based on an imminent hazard
- *Consent agreements* – negotiated agreements to avoid further penalty action, often involving significant costs beyond simple compliance
- *Public dissemination of information* – broadcasting the results of enforcement actions to increase deterrence

There are also many other external factors and players that influence companies' and their employees' actions.[112] Operators are subject to oversight by many different regulatory agencies. Insurance costs, the possibility of lawsuits from injured parties, harm to the company's reputation, and even the risk of whistleblower complaints all reinforce attention to compliance and good safety practices. At the same time, the evolutionary process of natural selection tends to weed out those who are not good at maximizing profit and shareholder value.[113]

It is widely acknowledged that performance-based regulations are more difficult to enforce, and this point is generally corroborated in the IM program by observations from inspectors, region directors, and program managers. Unless a company simply doesn't have something that it should have (e.g., a plan to determine valve placement), enforcement can involve considerable judgment and shades of gray. It's often more difficult in these cases to develop a preponderance of evidence to support a violation.

There are several issues that affect the effectiveness of enforcement actions:

- *A sense of moral obligation* might be one of the most important motivations for individuals to comply with well-founded requirements.
- *External factors*—like private liability for damages—add to the overall cost of non-compliance, and may reduce the "penalty" needed to incentivize compliance
- *Clear expectations, penalty amounts, certainty of punishment, probability of detection, and broad communication* are important for deterrence, particularly for companies that might have a lesser sense of moral obligation to comply.
- *Timeliness* is important for correcting unsafe conditions, and for connecting outcomes (penalties or costs) with actions (non-compliance).
- *A relatively simple and efficient process* is important to keep the process usable—and therefore *used*—by inspectors. It is also important in maximizing public benefits, as an enforcement program has a public cost.
- *Fairness* is important to establishing an efficient process, and in nurturing or supporting a sense of moral obligation and a strong safety culture.
- *A strong link to safety and environmental risk* is important to keep the enforcement program connected with the safety mission, and in reinforcing safety culture.

Timeliness and efficiency

The timeliness of civil penalty action has been a challenge for many years. From 2002 through the end of 2011, it has taken an average of *3 years and 7 months* to process an IM violation from the date the inspection ended until the final order was issued:

- An average of 10 months from the date the inspection ended to the time a notice of probable violation is issued to the operator, and
- 33 months to adjudicate the violation, from the time the NOPV was issued to the issuance of a final order.

Enforcement Processing Time (IM violations 2002-2011)

Year case opened	Number of citations	Average # days from inspection start to end	Average # days from inspection to notice	Average # days from notice to final order
GT	1,095	34	396	1,149
2006	187	17	290	1,161
2007	371	53	299	1,340
2008	236	18	365	803
2009	228	24	704	674
2010	34	55	244	497
2011	39	57	350	242
HL	1,635	20	241	917
2002	53	2	89	459
2003	53	34	201	1,226
2004	614	10	224	1,016
2005	324	5	222	916
2006	150	8	548	769
2007	192	25	202	927
2008	82	24	176	321
2009	81	86	162	990
2010	14	10	279	287
2011	72	106	234	232
Grand Total:	2,730	25	303	979

This has gotten much shorter in recent years. PHMSA closed a record number (102) of IM enforcement cases in 2011, and had by then reduced the total time from a high of 5 years and 4 months (in 2007) to a total of only *8 months* in 2011—about 4 months to issue the notice of probable violation, and another 4 months to adjudicate the case.

The number of cases had also dropped by almost six-fold at the same time, greatly reducing the workload, but processing clearly has become much timelier.

Some comments from inspectors:

Enforcement takes way too long, it can take years—we can't process things fast enough to matter.

The NOPV used to be 2-3 pages, now it's 12-15 pages. This might increase the success rate, but the level of effort is so high and it takes so long that the process is less effective.

The [agency] holds cases to a [difficult standard to meet]; we often lose when it comes down to judgment.

While the calendar time for processing and completing a case has improved substantially, many[114] inspectors in the field believe the enforcement process has become too cumbersome. It is possible that inspection integration will introduce additional delays into the enforcement process, as these kinds of inspections can require several months to complete.

Clearly there can be a trade-off between the simplicity of the process and the soundness of the evidence needed to support enforcement action. But the impact for IM deficiencies seems to be less use of the enforcement process, which undermines deterrence.

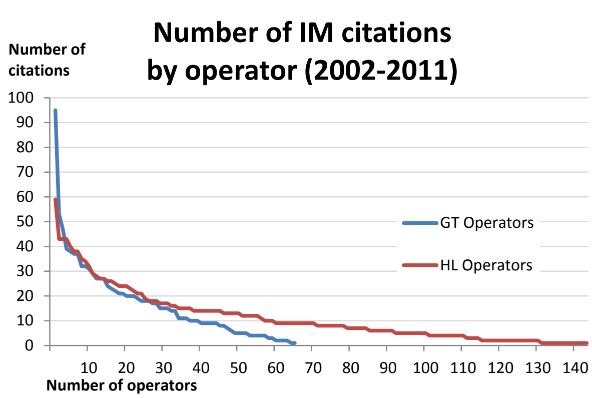
Deterrence

It seems fair to assume that most operators want to comply with the rules. Good safety programs can reduce the cost of accidents, and companies generally want to be perceived as good neighbors and good stewards. But there are many gray areas, particularly with performance-based requirements. Non-compliance can be a simple oversight, cutting corners, or even disagreeing with the relevance of a particular requirement.

Economic theory[115] suggests that the effectiveness of deterrence is a function of the *probability of detection* times the *consequences of detection*. This must be higher than the *benefit* (to the company) of non-compliance in order for a rational company to comply. And the threat must be clear and reasonably certain. As a practical matter, there are some significant data limitations that limit the agency's ability to estimate these variables. And the theory is actually more complicated, taking into account things like risk aversion. But it may be a useful, general framework for evaluating the data we have.

- For hazardous liquid systems, 143 operators (about 30% of the total number of operators) were cited for a total of 1,635 violations over 10 years of IM implementation (2002-2011). About half of these violations were referred for civil penalty action, and a total of $3,279,600 in penalties were assessed. Most of the other cases (about 45% of the total) resulted in notices of amendment.

- For gas transmission systems, 65 operators were cited for a total of 1,095 violations over six years of IM implementation (2006-2011). Only 21% of these violations were referred for civil penalty action, and a total of $1,114,500 in penalties were assessed. Most of the GT cases (73% of the total) resulted in notices of amendment.



Civil penalty actions were even more narrowly focused on a small number of operators. Over six years, only 14 GT operators have been assessed civil penalties—ranging from $17,500 to $306,000—for IM violations. Over 10 years, a total of 40 HL operators (fewer than 10%) have been assessed civil penalties, ranging from $5,000 to $526,000.

Notices of Amendment were the "preferred" way to resolve deficiencies or violations of the IM rules, particularly as "everyone was learning" during the first few years of IM implementation. Over half of all violations have been resolved this way—almost 80% for GT systems, which were subject to IM rules later. Civil penalty action (through a Notice of Probable Violation) was used for a little more than a third (37%) of all violations, compared to about 46% for violations of other parts of the pipeline safety code.

More striking, though, is a comparison of civil penalties assessed. Violations of the IM rules accounted for 26% of the citations from 2002-2011, but only 12% of the civil penalty amounts assessed.

The net effect: civil penalties were 2.6 times higher for non-IM violations than for IM violations. And civil penalty action is used even less now. Over the two-year period 2010-2011, NOPV action was used for only 22% of the citations—a decline of about 50% from previous years.

Enforcement Actions (IM violations 2002-2011)

# citations	Action					
Year case opened	Letter of Concern	Notice of Amendment	Notice of Probable Violation	Safety Order	Warning Letter	Grand Total
GT		**804**	**230**		**61**	**1,095**
2006		144	42		1	187
2007		231	137		3	371
2008		179	26		31	236
2009		189	19		20	228
2010		28	1		5	34
2011		33	5		1	39
HL	**1**	**737**	**792**	**1**	**104**	**1,635**
2002		22	30		1	53
2003	1	11	41			53
2004		155	456		3	614
2005		167	156		1	324
2006		91	41		18	150
2007		136	24		32	192
2008		56	4		22	82
2009		51	11		19	81
2010		10	3		1	14
2011		38	26	1	7	72
Grand Total	**1**	**1,541**	**1,022**	**1**	**165**	**2,730**
% of Total	0.04%	56.45%	37.44%	0.04%	6.04%	100.00%

This is consistent with the general idea that performance-based rules are harder to enforce. It is also consistent with an alternative explanation—that the focus has been more on correction that deterrence. Civil penalties, by contrast, are largely for deterrence.

These numbers, and the types of enforcement action, varied by region. The Western Region referred the largest number of violations, for both gas transmission and hazardous liquid operators, and the largest number of civil penalty actions (NOPVs). Based on interviews with field inspectors and Region Directors, these differences appear to reflect differences in enforcement approach more than operators' performance.

IM Enforcement Actions by Region (2002-2011)

Number of Citations	Region					
Enforcement actions	Central	Eastern	Southern	Southwest	Western	Grand Total
GT	**200**	**60**	**126**	**329**	**380**	**1,095**
Notice of Amendment	182	20	85	253	264	804
Notice of Probable Violation	6	39	28	74	83	230
Warning Letter	12	1	13	2	33	61
HL	**174**	**208**	**138**	**371**	**744**	**1,635**
Letter of Concern	1					1
Notice of Amendment	81	135	68	291	162	737
Notice of Probable Violation	92	70	64	70	496	792
Safety Order				1		1
Warning Letter		3	6	9	86	104
Grand Total	**374**	**268**	**264**	**700**	**1,124**	**2,730**

There is no direct way to assess whether these numbers are "good" or not with respect to deterrence. The average penalty per citation—for violations that are referred for civil penalty action—was $18,309. But the average civil penalty across all violations was only $1,610.

Is that sufficient to achieve deterrence? *Probably not*, for several reasons:

- Most violations do not result in any penalty. Operators know this from experience.
- Inspectors acknowledge that it's much harder to evaluate processes (like risk analysis, program evaluation, metrics, prevention and mitigation measures, reinspection intervals) under IM.
- When they find deficiencies, inspectors believe the enforcement process is overly cumbersome and time-consuming, so they don't use it as much as they might otherwise.
- Warnings, NOAs, and compliance orders tend to work toward getting something back into compliance. When used by themselves, these tend to undermine deterrence, as the only cost is getting something back into compliance.

The benefits (or costs avoided) for non-compliance and the probability of detection are not easily measured, so they are not used in making civil penalty recommendations or in determining civil penalty amounts in the adjudication process. Some inspectors have also pointed out that the program has a lot of leverage *after an accident*; much less during the course of routine inspections. This observation tends to support the argument that the agency might have more effect with a greater proportion of its "inspection" resources spent investigating accidents. In fact, corrective action orders and consent agreements—commonly imposing the greatest cost and obtaining the largest scope of company response—are typically used after an accident.

Influencing safety culture

Safety culture has been implicated in most really big accidents over the past 25 years, and there is a growing recognition that it is an essential element in achieving safety goals. Its importance is based on several key assumptions: the program is dealing with complex systems, where people can never really anticipate all the interactions that might lead to an accident; culture determines (to a large extent) how people act and make decisions; people are often the failure point, but also often the last safety mechanism before a failure gets out of control; and inspectors can't inspect everything.

These assumptions intersect some of the assumptions underlying the IM program and probably any performance-based regulatory system.

What a safety program really wants is for people to act safely – even when nobody is looking. This has several implications for enforcement. It suggests:

- Inspectors should be looking at good engineering practice, not just compliance; the program can't anticipate everything in its standards.
- Inspectors should overlook any non-compliance that results in a better safety decision. Sometimes acting safely means breaking the rules (the processes, procedures, or even the regulations), especially in an emergency.
- The program should do all this in a way that gets people to focus on what's really important for safety, not simply getting companies to spend enormous effort and resources on safety; fairness is important here, so that companies recognize that the program will discriminate between what's important and what's not.
- A sense of personal culpability helps the program achieve its objectives; anything PHMSA can do to advance this is likely to be helpful. This might mean simply identifying (to company management) individuals who failed to do what they were supposed to do.
- Organizational learning is important for safety culture. The program should encourage people in the regulated community to share their failures and even non-compliance if this can collectively make the system better as a result.

The program is doing many of these things in an informal way now, as inspectors understand intuitively that these things are important and exercise their discretion. But they are generally not yet part of, or central to, the policy guidance. In fact, this is an emerging area of interest and activity across many federal agencies.

The importance of safety culture also has several implications for shaping the broader enforcement program:

- We need to better understand the incentive structure in companies that drives behavior. This is, after all, one of the main competing demands, and one that is very much at play in making deterrence work.
- The regulatory framework needs to be cooperative and accommodative where companies are well-intentioned, to reinforce the moral appeal or sense of obligation and the need for experimentation.

- PHMSA should be backing up a focus on safety culture with a strong enforcement program where the consequences times the probability of detection is greater than the benefit of non-compliance.
- The agency itself routinely should be analyzing what works and what doesn't, based on a diagnostic assessment of deficiencies found and good practices observed during inspections and investigations, and feeding this back into the standards and rulemaking program as well as inspection guidance. PHMSA has done this to some extent for enforcement, but safety culture adds a new dimension.

Some general observations on enforcement

The evidence suggests that there has been a greater focus on corrective action than on deterrence when IM deficiencies were discovered. This might reflect a natural emphasis as any new program is implemented. It might also reflect a common understanding that most companies in this industry intended to comply with the rules, and that harsher enforcement actions might have had unintended effects. To some extent, it also reflects some lack of confidence in the enforcement process—tied particularly to perceived difficulty in using the process—by the inspectors who would use it.

A growing emphasis on safety culture suggests taking a broader approach to influencing behavior than the more traditional model of regulatory enforcement. This seems to be a work in progress, with many informal efforts already aiming to affect pipeline operators' actions "beyond compliance."[116] The Secretary's more recent guidance on Safety Management Systems, NTSB's recommendations highlighting safety culture, and many other interagency efforts to understand and advance safety culture, are reinforcing this general thrust.

Recently-issued enforcement guidance for the IM program was intended to help inspectors and States enforce the performance-based rules and clarify enforcement authorities, and to help the regulated industry understand the agency's expectations.

As the IM program matures, continuing to move to a more nuanced enforcement posture probably requires better data, more analysis, a more streamlined process, and better understanding of the process and standards for proving violations. This shift would also benefit from expertise in behavioral economics, behavioral psychology, and sociology to help explore how best to influence behavior in the companies the program regulates.

Program management

IM **marked a significant shift in thinking** about how pipeline operators would manage their systems. Performance-oriented standards were added to the body of pipeline safety regulations, bringing new requirements to assess and repair the physical infrastructure, evaluate risks, take actions to reduce or mitigate those risks, and put in place systems for continuous improvement. The IM rules aimed to put the burden squarely on the operator to manage its risks, and provided flexibility to do so.

Most people would probably agree that any program that focuses responsibility for safety on the organization directly involved has to be a good thing.[117] And you can't hold someone accountable for outcomes unless you provide flexibility in how to achieve those outcomes. The quandary is that that most of the expected safety outcomes are not evident in the data. At the same time, there are several gaps or complications in program design and implementation. And of course both of these are of interest to program management.

Developing and managing change

Several inspectors have raised concern about the agency's reliance on industry standards in the way IM was developed and implemented. They point to assessment methods, repair criteria, reassessment intervals, generally ambiguous language in the rules, and different requirements and repair criteria for gas vs. liquid pipelines as simply adopted from industry standards with no analytical basis. In fact, the preamble to the final rule dismissed some objections to the rule without explanation. And people who were involved in program development acknowledge the significance of industry standards in the final shape of the rules. In fact, these concerns reflect a deeper current in the regulatory field …

Regulators in general tend to become dependent on the regulated organizations to help them acquire and interpret information, which can undermine the regulatory process in a number of ways.[118] OMB policy[119] requires deference to industry standards, which can simply reflect the least common denominator, an approach that people can agree on. Industry research and consensus standards, cost estimates, and risk assessments can be very challenging to validate or refute with limited resources. Data collection is substantially constrained by the Paperwork Reduction Act. For pipeline safety, this greatly limits the data available for identifying important risk factors and for monitoring effective implementation of the program. And all of these issues are amplified with most performance-based approaches to regulation. Inspectors confront hundreds or thousands of different approaches and engineering evaluations, and the burden is on the regulator to prove that something *isn't* reasonable or technically justifiable.

Performance-based requirements can work if there are clear measures to judge performance, but there are not. One comment on the proposed rule[120] suggested that PHMSA couldn't evaluate the adequacy of operators' programs without specific requirements; the agency's response was simply that "OPS believes" performance-based language will best achieve effective IM programs, without further explanation. In implementation, the inspection protocols were filled with general qualifiers (like sufficient, adequate, reasonable, and technically-justifiable) about processes that inspectors don't have the time, experience, and specialized expertise to judge.

There is no standard for acceptable risk. The IM rules generally aim for standard processes, not outcomes. Apart from the repair criteria, there are no absolute standards. In fact, the guidance explicitly permits each operator to apply their own weights or values to the risk factors in their models.[121] Each operator is expected simply to prioritize its own actions and continually make improvements, but the public cannot expect even levels of risk. In fact, there appear to be large disparities in risk across different systems; the mere fact that past incidents are predictive of future incidents[122] suggests that operators are managing their systems to different standards of risk.

To some extent, this is a function of the legislative history of the program. Also, it follows a long tradition of acknowledging that each system has its own idiosyncrasies and must be managed with a view to those unique circumstances. And performance-based rules inherently provide that kind of flexibility. But every kind of risk targeting suggests that there is some higher, common standard of risk for comparison. It seems worth considering whether a common and absolute (quantitative) risk standard might be useful for managing the program more effectively.

Many of the problems in IM design and implementation identified in this evaluation were highlighted in public comments on the proposed rule, published April 24, 2000. The agency evidently misunderstood those comments or their significance. It was certainly a complicated rulemaking, charting a new direction, under considerable pressure to get a rule published. But this also reinforces a point made repeatedly in this evaluation—the agency did not, and does not, have the expertise in many specialized areas to design, implement, or manage a program like this. This might be a high priority gap to address as the agency works to improve the program.

How to close the loop – getting clear and convincing evidence of alignment between good intentions and real, on-the-ground delivery?[123] In one common formulation (Plan – Do – Check – Act), two critical elements are often missing: checking and acting. These require good performance measures, investigations, audits, records and reporting, management reviews, corrective actions, and process improvement.

We don't really know what would have happened without integrity management. But if it is important to know, the program could build program evaluation into the design and implementation of a program to help answer this kind of question. In fact, OPS (like many other federal safety programs) doesn't do this with any of its programs. It pilot tests many programs, including IM and integrated inspections, but these are largely for the purpose of working out the logistics of implementation, forms and processes. Without a strong analytical foundation, testing of assumptions, well thought-out metrics, and evaluation plans built into program design, change can be over-influenced by assumptions and beliefs. These are more commonly seen in fields like public health. They are technically feasible, but challenging to build.

Dealing with "normal," organizational accidents

Over the past several decades, PHMSA and the pipeline industry have identified many systemic problems in pipeline systems based on failure mode—for example, problems with low frequency electrical-resistance welded (LF-ERW) pipe, casings, third-party damage, corrosion, etc.—and have largely worked them out of the system. The long term trends in safety outcomes, particularly a decline in incidents with death or major injury, reflect these advances. But as problems are addressed, it has become increasingly difficult to identify these kinds of risk concentration to pursue.

IM was aimed at a deeper level of risk. Some people talk about *low-probability high-consequence* (LPHC) accidents to describe where these risks fall in the spectrum of likelihood vs. effect. Others elaborate on this special kind of risk.

- OSHA distinguishes *process safety* (things like explosions or releases of toxic vapors that create a much larger hazard) from more localized things like trips, falls, and burns (*occupational safety*). The Baker Panel report on the BP refinery explosion in Texas City provides an extensive discussion of challenges and potential solutions for managing process safety.

- James Reason has focused attention more on latent conditions and the mechanisms of failure. In *Managing the Risks of Organizational Accidents*, he observes that major accidents commonly arise from the unforeseen interactions of human and organizational factors, leading to what he calls an "*organizational accident*."[124]

- Charles Perrow suggests that there is an inherent risk in systems that are both complex and tightly-coupled—leading to what he characterizes as "*normal accidents*." [125] His book (with the same title) begins with a discussion of the Three Mile Island accident, and traces the sociological issues associated with complex technologies.

Natural gas and hazardous liquid pipelines seem to meet all these criteria. These are complicated systems, operating in difficult and sometimes hostile environments, and requiring designs with many interactions that often are not visible. Most of the pipe is buried. Since no operation is perfect, there will be failures ("normal accidents"). And the potential for a system accident with high consequences can increase in a poorly-run organization, as there are more possible failures to interact in unexpected ways.

Organizational accidents present a real challenge for the program. PHMSA doesn't have data on these kinds of failures because it doesn't have a good conceptual model for them. The program investigates very few of them. The risk models don't address them at all. And PHMSA's data collection doesn't address them at all. The IM program provides new information on pipeline integrity. But in some respects, the program has made a high-risk system increasingly complex and more tightly coupled, with smaller margins for safety.

The literature provides a number of ways to mitigate this kind of hazard:

- Tracking precursor failures—sometimes called "near misses"—can provide a leading indicator to help identify riskier systems or companies for preventive action.
- Adding barriers—the airline industry, for example, has reduced coupling and complexity with extensive backup systems and possibilities for de-coupling. After setting what some believed to be an unrealistic goal of an 80% reduction in commercial aircraft accidents, the agency actually achieved its goal.
- Building a strong safety culture, safety management system (SMS), and using the principles for running a high-reliability organization (HRO) can provide additional checks, backups, and continuous learning that can reduce the risk of system failures.[126]

You can't really fix all these things without re-tooling the program, reallocating substantial resources to it, focusing first on building more intelligence into it. The program needs "a robust information system to drive decision making with data." It is nowhere near that today.

Implementing SMS

In May 2012, the Secretary of Transportation provided guidance on developing safety management systems, and directed the heads of all operating administrations to conduct a gap analysis and develop an implementation plan for its own operations. The Secretary also suggested encouraging the regulated industry to implement SMS on a voluntary basis.

The Secretary's guidance generally includes the common elements of performance-based regulations and safety management systems described in Program Design. It adds one new element, not explicitly included in most the other systems: a strong safety culture.

For DOT, SMS will generally include four major elements, each with 5-10 specific features, tailored to any special circumstances that might be important in different modes of transportation:

Safety Policy

- *Describes what the organization is trying to achieve* through its SMS;
- *Outlines the requirements, methods, and processes* the organization will use to achieve the desired safety outcomes;
- *Establishes senior management's commitment* and expectation that the organization will incorporate and continually improve safety. The safety policy further establishes and defines senior management's expectation of high safety performance;
- *Reflects management's commitment to implementing procedures and processes* for establishing and meeting measurable and attainable safety objectives, and supports promotion of a positive safety culture;
- *Establishes roles, responsibilities, and accountabilities* regarding the organization's safety performance; and
- *Outlines an emergency response plan* that provides for the safe transition between normal and emergency operations where applicable.

Safety Risk Management

- *Describes the system of interest*: Establishes an understanding of critical system design and performance factors, processes, and activities to identify hazards
- *Identifies hazards*: Identifies and documents hazards or those things that could go wrong in sufficient detail to determine associated safety risks (within the system description)
- *Analyzes safety risk*: Determines and analyzes the severity and likelihood of potential events associated with identified hazards
- *Assesses safety risk*: Compares the safety risk of each identified hazard to established safety performance targets and/or ranks hazards based on risk
- *Controls safety risk*: Designs and implements safety risk control(s) for hazards with associated unacceptable risk

Safety Assurance

- *Data/information acquisition*. Collect, manage, and monitor operational data to assess operational system and SMS performance, identify new hazards, and measure the effectiveness of safety risk controls.
- *Reporting system*. Establish and maintain a safety reporting system in which stakeholders can report safety issues or concerns. Data obtained from this system are monitored to identify emerging hazards and to assess performance of risk controls in the operational systems.
- *Investigation*. Collect data and investigate incidents and accidents to identify new hazards or ineffective safety risk controls.
- *Monitoring, Evaluations and Audits*. Monitor, evaluate, or audit standards, systems, programs, and processes on a routine basis to determine the performance and effectiveness of safety risk controls. Also conduct regularly scheduled evaluations of the SMS to determine if the SMS as a whole conforms to its requirements.
- *Data/information analys*is. Analyze data to assess safety performance, identify new hazards, and measure the effectiveness of safety risk controls
- *System assessment*. Conduct assessments of the effectiveness of safety risk controls and overall SMS performance
- *Corrective action*. Prioritize and implement corrective actions to mitigate or eliminate problems identified during system assessments
- *Management reviews*. Conduct regular reviews of SMS effectiveness and assess the need for changes to the SMS

Safety Promotion (the ten most critical elements of a strong safety culture)

- *Leadership is clearly committed to safety*
- *There is open and effective communication across the organization*
- *Employees feel personally responsible for safety*
- *The organization practices continuous learning*
- *There is a safety-conscious work environment*
- *Reporting systems are clearly defined and non-punitive*
- *Decisions demonstrate that safety is prioritized over competing demands*
- *Mutual trust is fostered between employees and the organization*

- *The organization is fair and consistent in responding to safety concerns*
- *Training and resources are available to support safety*

Integrity management is an SMS type of system. It does not include every element outlined in the Secretary's guidance, but it includes many, and those parts are mandatory. Many other elements (training, qualification, operating procedures, emergency response) of an SMS are addressed to some extent in other parts of the pipeline safety code, but not integrated into the overall program of risk evaluation/management under IM.

Applying SMS principles to program operation might address many of the challenges outlined in this evaluation. This is an idea the Associate Administrator has talked about for some time—applying IM (or SMS) principles to the Office of Pipeline Safety itself—and it is exactly what the Secretary is now leading and supporting. It could help clarify leadership support, build a stronger analytical capability, re-work data collection programs, establish a more far-reaching system for reporting failures and near misses, build a much stronger investigations program, regularize and institutionalize management reviews, and build a stronger safety culture.

Safety culture is widely-acknowledged to be important in achieving safety results for large and complex systems. But all of the research on this has focused on the culture of an *operating* organization—where people are doing production-oriented things that have intrinsic risk. These are generally companies, like pipeline operators, or operational programs in agencies, like air traffic control, where an action or omission can lead to immediate risk.

Regulators have an important role in encouraging and supporting a strong safety culture in the companies they regulate. There are many areas where the regulator's actions could either support *or discourage* a strong safety culture.[127] And there is an argument that regulators should lead by example. But it's not clear that safety culture means the same thing when talking about a company and a regulator that oversees a company. Many of the questions in existing safety culture surveys don't seem to apply to a federal safety agency.

PHMSA should be thinking about how to measure these things in the companies it regulates. There *is* a scientific basis for doing that, with survey scales supported by research. And the agency might further explore how its actions affect safety culture in the regulated community.

Ongoing work and current plans

PHMSA has underway several initiatives to improve its data, risk modeling, performance measures, accident investigations, enforcement policy. Also, and more directly relevant, it is considering a range of options for changes in the integrity management program based on experience so far. The agency has organized meetings with stakeholders and the public to invite comment on PBRs and metrics, published safety advisories to operators about data quality and metrics, and published Advance Notices of Proposed Rulemaking (ANPRMs)[128] for both hazardous liquid and gas transmission pipelines. These ANPRMs solicit public comment on several IM requirements, including:

- Modifying the definition of an HCA, and/or extending some IM requirements beyond HCAs (GT & HL)
- Making requirements related to the nature and application of risk models more prescriptive (GT)
- Strengthening requirements on selection and use of assessment methods (GT)
- Adopting standards for ILI tool performance (GT & HL)
- Prescribing methods for validating ILI tool performance (GT)
- Adopting qualification standards for people interpreting ILI data (GT & HL)
- Revising requirements for collecting, validating, and integrating pipeline data (GT)
- Modifying repair criteria (GT & HL)
- Strengthening IM requirements for preventive and mitigative measures (GT)
- Strengthening requirements for applying knowledge gained through the IM program (GT)

The ANPRM's also ask for comment on strengthening or expanding certain non-IM requirements associated with system integrity, including:

- Removing exceptions and exemptions for certain pipelines (GT & HL)
- Corrosion control, including methods for preventing, detecting, assessing and remediating stress corrosion cracking (GT & HL)
- Expanding leak detection requirements (HL)
- Specifying valve spacing and the need for remotely-controlled valves or emergency flow-restricting devices (GT & HL)
- Addressing pipe with longitudinal weld seams with systemic integrity issues (GT)
- Establishing requirements for underground storage facilities (GT & HL)
- Addressing quality management systems and management of change (GT)

This evaluation provides some additional analytical background that could help choose, shape, or add to these ideas for program improvement.

A framework for improving the IM program

The changes envisioned in the ANPRMs might address many of the technical issues identified in this evaluation. But they are all focused on what the regulated industry should do. If these are not coupled with some fundamental changes (outlined below) in PHMSA's oversight, I believe there is a substantial risk that some of these efforts will be misguided and/or disappointing in their effects.

In making broad changes in the IM program, expanding the use of integrated inspections, and trying to implement SMS, there are a few things that could substantially increase the probability of success:

- *Building much more substantial expertise in the areas need to oversee performance-based regulations*—to broadly re-tool the IM rules and the inspection and enforcement processes in a way that more clearly considers how the program influences behavior.
- *Building a much stronger analytical capability*—including expertise in social science, economics, risk evaluation, statistics, and program evaluation.
- *Developing a good conceptual model of failures*—to underpin PHMSA's data collection, inspection, investigations, and rulemaking programs.
- *Building a robust information system*—including overhauling the data collection program and metrics, starting with the question: *What do we need to know?*
- *Correcting the rules and guidance on risk factors and risk modeling*—to fix known problems that are undermining effective safety decisions by operators. *Re-shaping the inspection program into a more forensic, investigative approach*—to adapt to the special challenges in performance-based regulation.
- *Expanding the accident investigations program*—to provide the primary feedback loop into the program for learning about failures, assessing program effectiveness, and redirecting effort.
- *Developing a system for managing change*—grounded in credible analysis, testable assumptions, and broad input; and with evaluation built into the design.

Many of these will require a substantial investment of resources, either new or (where possible) redirected from other activities.

Appendix A - Common Acronyms

API	American Petroleum Institute
ASME	American Society of Mechanical Engineers
BSEE	Bureau of Safety and Environmental Enforcement (Department of Interior)
DA	Direct assessment
DOT	Department of Transportation
ECDA	External corrosion direct assessment
EPA	Environmental Protection Agency
GAO	Government Accountability Office
GT	Gas transmission
HCA	High consequence area (including a liquid pipeline segment that *could affect* an HCA)
HL	Hazardous liquid
HRO	High reliability organization
HSE	Health and Safety Executive (UK)
II	Integrated inspection
ILI	In line inspection
IM	Integrity management
ISM	International Safety Management Code (applies to maritime shipping)
LPHC	Low probability high consequence
MAOP	Maximum allowable operating pressure (gas systems)
MOP	Maximum operating pressure (liquid systems)
NTSB	National Transportation Safety Board
OIG	Office of Inspector General (Department of Transportation)
OPS	Office of Pipeline Safety (within PHMSA)
OSHA	Occupational Safety and Health Administration (Department of Labor)
PBR	Performance-based regulation
PHMSA	Pipeline and Hazardous Materials Safety Administration (Dept. of Transportation)
PSM	Process safety management (OSHA's performance-based regulation)
RMP	Risk management program (EPA's performance-based regulation)
RRIM	Risk ranking index model, used to target PHMSA's integrated inspections
RSPA	Research and Special Programs Administration, the predecessor to PHMSA
SEMS	Safety & environmental management system (BSEE's performance-based regulation)
SMS	Safety management system
TRB	Transportation Research Board
USCG	U.S. Coast Guard (Department of Homeland Security)

Appendix B - Methodology & Limitations

Evidence and the evaluation process

The evaluation used interviews; analysis of data and program documentation; program logic modeling; development of a program theory of change; and review of legislation, rulemaking, past evaluations, investigation reports, risk evaluation literature, external standards, and other agencies' practices—all to develop the factual basis for understanding the current program and ideas for improving it.

Over the course of the evaluation, interviews were conducted with 15 senior inspectors, field supervisors, and region directors; 7 programs managers and technical advisors involved in program development and implementation; 4 technical specialists in pipeline data and analysis; one enforcement attorney; several current and former employees of pipeline companies; one representative of a public interest group; and 4 other federal regulators of the oil and gas industry.

During the evaluation, OPS jointly (with five other agencies) sponsored a meeting in Texas City to explore issues in performance-based regulation of the oil and gas industry. I attended that two-day meeting and continued some of the discussions with other agencies afterward.

Data were analyzed from PHMSA (operator incident reports, system characteristics and IM metrics in operators' annual reports, inspection and enforcement data); Census Bureau (population, housing starts); and the Energy Information Administration (energy production, consumption, and many related variables used in the National Energy Modeling System). The results of these analyses were reviewed by senior data/analysis specialists in PHMSA.

As much as possible, elements of the program and outcomes were evaluated against established criteria—including legislation and published rules, regulatory evaluations, DOT's Information Quality Guidelines, standards, and the risk literature. Safety outcomes were evaluated with the assistance of a senior statistician (specializing in forecasting and time series analysis) in the Bureau of Transportation Statistics.

The evaluation generally followed the Program Evaluation Standards published by the Joint Committee on Standards for Educational Evaluation (JCSEE).

Analysis of the accident/incident trends

The trend data on pages 24-32—and particularly the chart on page 32—were drawn from DOT/PHMSA incident data as reported by operators under 49 CFR 191.3 (for natural gas transmission pipelines) and 49 CFR 195.50 (for hazardous liquid pipelines). In general, these regulations require reports whenever there is a release resulting in death, injury requiring hospitalization, or property damage of $50,000 or more. Both gas and liquid systems require reporting of certain additional failures based on the amount released.

Normalizing the data

For each sector (gas and liquid), I normalized the data to help make more reliable comparisons across time. This is essentially the same methodology the program has used to track "significant incidents" on the PHMSA website.

- I adjusted property damage estimates for inflation (using the GDP deflator), and omitted incidents with property damage below $50,000 in 1986 dollars unless the incident was required to be reported under at least one other reporting criterion.
- I omitted liquid spills of less than 50 barrels (2,100 gallons), since the reporting criteria changed (to 5 gallons) in 2002, breaking the time series without this adjustment.
- I omitted incidents that did not meet any of the reporting criteria (these might have been submitted as significant in the judgment of the operator, but with no objective criteria to ensure consistency).

For natural gas systems, I omitted incidents from gas gathering systems, as these are not covered under the IM rules.

Breaking out HCA vs. non-HCA incidents

For each sector, I analyzed the long-term trends (1986-2012) as well as a more detailed breakout of the post-IM trends (HCA vs. non-HCA). There are no HCA data before 2002.

For hazardous liquid systems, I considered 2001 to be a transition year. The IM rule for liquid systems was published December 1, 2000 (for systems with 500 miles or more, initially), and the rules were effective March 31, 2001. Operators were required to identify segments that could affect HCAs by December 31, 2001, and to have a written IM program by March 31, 2002. Operators were also required to begin reporting IM data in 2002. So I used 2002 as the first year of implementation in the analysis. It was not an exact starting point, but appeared to be reasonable and useful for distinguishing pre-IM incidents (1986-2000) from post-IM incidents (from 2002 forward).

For gas transmission systems, I considered 2004 to be a transition year. The final rule was published December 15, 2003 and the rules were effective January 14, 2004. Operators were required to identify pipeline segments in HCAs and to develop and follow a written IM program by December 17, 2004. So I used 2005 as the first year of implementation in the analysis.

For both sectors, I used the "HCA" data field in operators' reports to distinguish accidents or incidents in HCAs (or in segments that could affect an HCA) from those not in HCAs. Some analysts have pointed out that different companies might have been interpreting this data field differently, particularly for liquid pipelines where a segment *could affect* an HCA without being in an HCA. There is no obvious way to improve the historical data at this point, but if the differences were company-specific then it seems reasonable to assume that these differences would not substantially affect industry-wide comparisons over time.

Some liquid pipeline incidents were reported with "N/A" or "blank" HCA data. But these were few in number (10 total over the 11-year period 2002-2012), and none involved death or major injury. The property damage for these incidents was less than 2/100 of 1% of the total.

Breaking out consequences

For each sector, after normalizing the data, I broke out the consequences that were described in the section on expected results. This included:

- Number of reported incidents
- Number of high consequence incidents
- Number of deaths
- Number of injuries
- Amount of property damage (in 2012 dollars)
- Number of incidents attributed to corrosion and material failure
- Number of hazardous liquid spills with environmental impacts
- Barrels spilled from hazardous liquid pipelines

Reported incidents include significant incidents only, for comparability of the data over the long term (1986-2012).

High consequence incidents—for purposed of this evaluation—include any incident involving a death, major injury, property damage of $500,000 or more (in 2012 dollars), or a liquid release of 100,000 gallons or more (a "major" spill as defined in the Federal On Scene Coordinator's Guide to Environmental Response). These latter two criteria are each about 10% of the total number of reported incidents, with significant overlap.

This is not a standard category of incident used by the pipeline safety program. I assumed that any death or major injury would be considered a high consequence incident in most people's minds; the agency's primary performance measure focuses on these two consequences. I used $500,000 in property damage as an additional criterion because it was close to the monetized value of a statistical injury used in regulatory evaluations over the period of time the IM was rule was in effect; it was a simple, rounded value. I used 100,000 gallons spilled because this is the threshold used for decades by the Coast Guard and EPA in distinguishing a major spill from a medium spill. Overall, these criteria are simply providing a rough way to estimate the changes in higher-consequence incidents over time.

Deaths includes all reported fatalities, including workers as well as the general public.

The number of injuries was adjusted to remove 1,851 non-serious injuries in a single 1994 accident—reported before the criteria for injury reporting was clarified to mean injuries requiring hospitalization. This is a known anomaly in the data that was removed consistent with the general logic for normalizing the data in "significant incidents."

Property damage was converted from nominal dollars (after normalizing the data explained previously to correct for increasing levels of reporting simply due to inflation) to 2012 dollars to help ground the data in the present context.

Corrosion and material failure are causes that are particularly relevant for the IM program. Excavation damage was broken out as well because it showed it substantial trend that could help explain the overall results we are seeing in the data.

Spills with environmental impacts include any spill of 5 barrels or more where the operator has indicated any environmental consequences (impacting soil, water, fish, birds, or other terrestrial wildlife). The 5-barrel threshold reflects the reporting criteria for these impacts; below 5 barrels there is a shorter form for reporting incidents**.**

Barrels spilled reflects the total amount lost, and is not adjusted for product recovered. This is sometimes explicitly converted to gallons in the report (one barrel of hazardous liquid equals 42 U.S. gallons).

The data used in the section "The Data – *What actually happened*" (pp. 24-32, and particularly the summary graphics on page 32) are tabulated below.

HL Accidents and Consequences (1986-2012)

Significant accidents (for comparability across long term)

Year	# of accidents	# High Conseq. Incidents	# of fatalities	# of injuries	Property damage (2012 $)	Barrels lost	Spills w/env. impacts
1986	194	40	4	32	$ 29,095,673	282,611	
1987	215	39	3	20	$ 22,866,285	395,649	
1988	167	36	2	19	$ 54,994,761	198,111	
1989	136	29	3	38	$ 13,904,604	201,504	
1990	140	21	3	7	$ 24,218,481	123,827	
1991	166	29	0	9	$ 57,064,223	200,210	
1992	170	41	5	38	$ 57,206,209	136,769	
1993	154	30	0	10	$ 40,626,134	116,132	
1994	178	45	1	7	$ 87,566,517	163,920	
1995	156	24	3	11	$ 44,448,643	109,931	
1996	174	40	5	13	$ 116,717,603	160,188	
1997	162	34	0	5	$ 74,326,620	195,421	
1998	140	36	2	6	$ 84,068,808	149,348	
1999	147	38	4	20	$ 113,096,749	167,082	
2000	135	38	1	4	$ 194,263,049	108,614	
2001	108	27	0	10	$ 31,032,136	98,046	Transition year
2002	133	24	1	0	$ 60,052,289	95,664	153
2003	123	31	0	5	$ 77,989,581	80,032	149
2004	135	42	5	16	$ 193,389,148	88,211	138
2005	129	42	2	2	$ 348,338,407	137,052	127
2006	107	29	0	2	$ 79,918,489	136,500	106
2007	109	32	4	10	$ 62,378,882	94,083	97
2008	122	42	2	2	$ 152,300,842	101,057	128
2009	108	34	4	4	$ 72,548,809	50,463	111
2010	121	40	1	4	$ 1,059,852,912	174,101	94
2011	140	48	1	2	$ 246,328,538	138,216	117
2012	128	33	3	4	$ 104,061,417	53,172	123
Grand Total	**3897**	**944**	**59**	**2151**	**$ 3,502,655,810**	**3,955,914**	**1343**

HL Accidents and Consequences (1986-2012)

Significant accidents (for comparability across long term)

HCA: YES

Year	# of accidents	# High Conseq. Incidents	# of fatalities	# of injuries	Property damage (2012 $)	Barrels lost	Spills w/env. impacts
2002	38	10	0	0	$ 28,968,317	23,083	45
2003	49	18	0	5	$ 50,337,658	27,331	47
2004	44	17	5	15	$ 46,770,817	22,551	41
2005	52	20	0	2	$ 120,212,845	70,679	49
2006	44	13	0	0	$ 43,131,892	16,779	43
2007	47	11	0	2	$ 26,096,560	18,018	40
2008	63	29	1	1	$ 74,003,404	25,852	67
2009	45	14	1	3	$ 24,679,451	13,413	50
2010	48	17	0	1	$ 982,464,037	41,415	40
2011	58	22	0	0	$ 194,392,022	21,052	52
2012	50	13	2	3	$ 62,440,245	12,132	46
Grand Total	**538**	**184**	**9**	**32**	**1,653,497,249**	**292,306**	**520**

HCA: NO

Year	# of accidents	# High Conseq. Incidents	# of fatalities	# of injuries	Property damage (2012 $)	Barrels lost	Spills w/env. impacts
2002	89	13	1	0	$ 30,604,166	68,267	108
2003	73	13	0	0	$ 27,651,848	52,696	102
2004	91	25	0	1	$ 146,618,331	65,660	97
2005	76	22	2	0	$ 228,125,391	66,368	78
2006	63	16	0	2	$ 36,786,597	119,721	63
2007	60	21	4	8	$ 36,254,935	76,055	57
2008	59	13	1	1	$ 78,297,439	75,204	61
2009	63	20	3	1	$ 47,869,358	37,050	61
2010	73	23	1	3	$ 77,388,875	132,686	54
2011	82	26	1	2	$ 51,936,516	117,164	65
2012	78	20	1	1	$ 41,621,172	41,040	77
Grand Total	**807**	**212**	**14**	**19**	**803,154,629**	**851,911**	**823**

HCA N/A or blank

Year	# of accidents	# High Conseq. Incidents	# of fatalities	# of injuries	Property damage (2012 $)	Barrels lost	Spills w/env. impacts
2002	6	6	0	0	$ 479,806	4,314	0
2003	1	1	0	0	$ 74	5	0
2005	1	1	0	0	$ 171	5	0
2007	2	2	0	0	$ 27,387	10	0
Grand Total	**10**	**10**	**0**	**0**	**507,438**	**4,334**	**0**

HL Accidents by Cause (1986-2012)

Significant accidents (for comparability across long term)

Year	Corrosion	Excavation damage	Natural force damage	Human error	Material failure	All other causes	Other outside force damage	Grand Total
1986	47	65	4	10	21	*47*		194
1987	57	59	3	10	27	*59*		215
1988	45	50	1	12	20	*39*		167
1989	27	41	3	11	23	*31*		136
1990	27	34	2	10	26	*41*		140
1991	47	44	3	11	31	*30*		166
1992	34	49	1	11	29	*46*		170
1993	28	47	5	11	24	*39*		154
1994	30	30	12	9	43	*54*		178
1995	32	35	10	24	22	*33*		156
1996	53	42	2	10	23	*44*		174
1997	46	36	4	10	20	*46*		162
1998	34	32	7	6	22	*39*		140
1999	24	25	2	14	22	*60*		147
2000	31	34		9	21	*40*		135
2001	33	21	3	9	6	*36*		108
2002	37	23	5	8	44	11	5	133
2003	31	18	7	8	42	12	5	123
2004	37	23	17	4	39	9	6	135
2005	28	13	17	9	47	10	5	129
2006	32	13	5	8	31	12	6	107
2007	28	16	8	10	35	7	5	109
2008	33	18	6	17	39	7	2	122
2009	23	15	9	13	38	7	3	108
2010	26	13	8	14	50	5	5	121
2011	39	17	5	12	57	5	5	140
2012	32	8	4	12	56	13	3	128
Grand Total	**941**	**821**	**153**	**292**	**858**	**782**	**50**	**3897**

GT Incidents and Consequences (1986-2012)

Significant accidents (for comparability across long term)
Transmission systems only; not including gas gathering systems

Year	# of accidents	# High Conseq. Incidents	# of fatalities	# of injuries	Property damage (2012 $)
1986	40	13	3	13	$ 18,344,586
1987	23	9	0	15	$ 5,230,885
1988	39	9	2	9	$ 14,667,155
1989	40	18	22	28	$ 29,554,251
1990	46	19	0	17	$ 16,315,529
1991	29	9	0	11	$ 16,298,463
1992	42	18	3	14	$ 34,084,019
1993	50	23	1	16	$ 31,708,275
1994	45	18	0	22	$ 61,676,197
1995	32	14	2	7	$ 13,061,909
1996	43	9	1	5	$ 16,558,109
1997	33	8	1	5	$ 14,249,596
1998	51	19	1	11	$ 53,986,798
1999	37	10	2	8	$ 20,061,757
2000	54	14	15	16	$ 20,792,502
2001	50	12	2	5	$ 16,144,347
2002	56	15	1	4	$ 30,666,863
2003	70	25	1	8	$ 57,954,719
2004	62	12	0	2	$ 39,893,236
2005	111	50	0	5	$ 338,960,203
2006	78	18	3	3	$ 42,424,599
2007	75	27	2	7	$ 65,035,243
2008	73	37	0	5	$ 267,775,489
2009	73	26	0	11	$ 56,878,892
2010	79	24	10	61	$ 420,342,766
2011	82	29	0	1	$ 96,553,067
2012	62	20	0	7	$ 51,531,351
Grand Total	**1475**	**505**	**72**	**316**	**$ 1,850,750,806**

GT Incidents and Consequences (1986-2012)

Significant accidents (for comparability across long term)
Transmission systems only; not including gas gathering systems

HCA: YES

Year	# of incidents	# High Conseq. Incidents	# of fatalities	# of injuries	Property damage (2012 $)
2004	1	0	0	0	$ 99,695
2005	3	0	0	0	$ 515,441
2006	3	0	0	0	$ 896,933
2007	6	1	0	0	$ 1,881,840
2008	2	1	0	1	$ 178,932
2009	4	2	0	0	$ 8,246,251
2010	5	1	8	51	$ 387,723,208
2011	7	1	0	0	$ 6,466,214
2012	12	4	0	0	$ 5,847,864
Grand Total	**43**	**10**	**8**	**52**	**$ 411,856,379**

HCA: NO

Year	# of incidents	# High Conseq. Incidents	# of fatalities	# of injuries	Property damage (2012 $)
2004	61	12	0	2	$ 39,793,541
2005	108	50	0	5	$ 338,444,762
2006	75	18	3	3	$ 41,527,666
2007	69	26	2	7	$ 63,153,403
2008	71	36	0	4	$ 267,596,557
2009	69	24	0	11	$ 48,632,641
2010	74	23	2	10	$ 32,619,557
2011	75	28	0	1	$ 90,086,853
2012	50	16	0	7	$ 45,683,487
Grand Total	**652**	**233**	**7**	**50**	**$ 967,538,466**

GT Incidents by Cause (1986-2012)

Significant accidents (for comparability across long term)
Transmission systems only; not including gas gathering systems

Year	Corrosion	Excavation damage	Natural force damage	Human error	Material failure	All other causes	Other outside force damage	Grand Total
1986	6	8	1		11	*14*		40
1987	6	7			1	*9*		23
1988	6	12	4		6	*11*		39
1989	7	15			5	*13*		40
1990	10	14	6		8	*8*		46
1991	7	12	2		2	*6*		29
1992	6	12	5		5	*14*		42
1993	10	6	6		9	*19*		50
1994	16	12	4		4	*9*		45
1995	2	8	6		7	*9*		32
1996	8	19	1		6	*9*		43
1997	9	11	4		3	*6*		33
1998	13	12	3		11	*12*		51
1999	7	11	3		6	*10*		37
2000	21	11	3		3	*16*		54
2001	10	15	3		8	*14*		50
2002	17	7	5	1	17	5	4	56
2003	16	11	2	6	19	9	7	70
2004	18	13	7	1	14	4	5	62
2005	15	10	43	4	21	9	9	111
2006	15	9	2	3	26	16	7	78
2007	30	10	3	2	13	13	4	75
2008	13	8	23	3	13	7	6	73
2009	14	5	8	1	23	11	11	73
2010	26	10	3	3	24	9	4	79
2011	17	8	13	6	27	6	5	82
2012	21	5	2	2	21	8	3	62
Grand Total	**346**	**281**	**162**	**32**	**313**	**276**	**65**	**1475**

GT Incidents by Cause and HCA Location (1986-2012)

Significant accidents (for comparability across long term)
Transmission systems only; not including gas gathering systems

HCA: YES

Year	Corrosion	Excavation damage	Natural force damage	Human error	Material failure	All other causes	Other outside force damage	Grand Total
2004					1			1
2005		2	1					3
2006		1		1		1		3
2007		2	1		2	1		6
2008		1			1			2
2009		1				1	2	4
2010	1	1			2		1	5
2011	1	2	1	1	1	1		7
2012		1	1		6	4		12
Grand Total	**2**	**11**	**4**	**2**	**13**	**8**	**3**	**43**

HCA: NO

Year	Corrosion	Excavation damage	Natural force damage	Human error	Material failure	All other causes	Other outside force damage	Grand Total
2004	18	13	7	1	13	4	5	61
2005	15	8	42	4	21	9	9	108
2006	15	8	2	2	26	15	7	75
2007	30	8	2	2	11	12	4	69
2008	13	7	23	3	12	7	6	71
2009	14	4	8	1	23	10	9	69
2010	25	9	3	3	22	9	3	74
2011	16	6	12	5	26	5	5	75
2012	21	4	1	2	15	4	3	50
Grand Total	**167**	**67**	**100**	**23**	**169**	**75**	**51**	**652**

Evaluating the trends pre- and post-IM

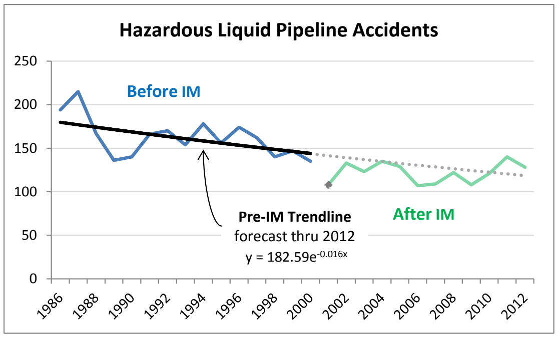
For each sector, and for each of the incident consequences that were analyzed, I plotted the data from 1986-2000 (for liquid pipelines) and 1986-2004 (for gas transmission) to provide the pre-IM baseline. For each of these time series, I ran an exponential regression on the data to help identify the trend over time. An exponential regression is useful for representing diminishing (or increasing) returns—logically consistent with the kinds of trends we expect to see with time series data. This established the starting point—a reflection of the *underlying risk* at the beginning of IM implementation, apart from annual variation in the data—for comparison. This method accounts for any trend better than taking averages of multiple years.

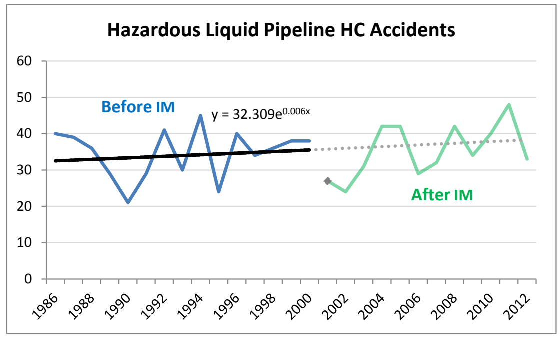
While the regulatory evaluations generally expected decreases in these outcomes, the pre-IM data showed both some rising and some declining trends. The usual approach in developing economic analyses for rulemaking is to compare the expected impacts from the rule to a "baseline." The baseline is essentially what is expected to happen without the rule. Commonly, this is a forecast from previous trends, which have already captured all the programs and external factors acting on the systems. For this evaluation, I simply projected the pre-IM trendlines as a basis for an initial comparison. Post-IM years were compared to this regression line to evaluate whether there was an increase, decrease, or no real change in the data.

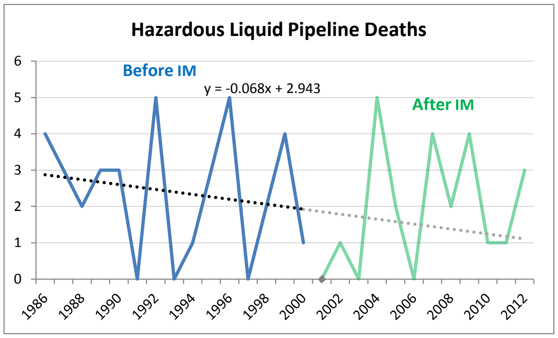
The patterns were fairly obvious. In many cases, all of the values after IM implementation (from 2002-2012 for HL systems, and from 2005-2012 for GT systems) were above the baseline, or all (but one) were below it. These cases were judged to be either in the "wrong" direction or the "right" direction, respectively. Two cases were judged slightly differently from the rest:

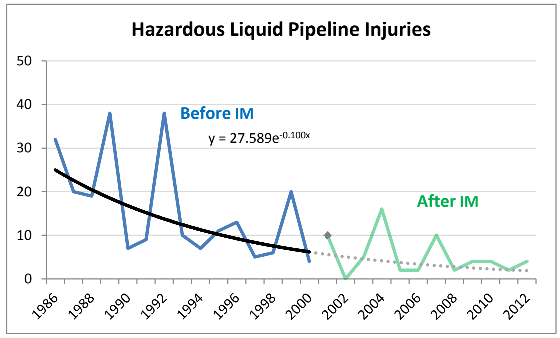
- For HL accidents caused by material failure, all post-IM values were above the baseline, but there was a change in reporting in 2002 that could have skewed the data. This was not judged to be the "wrong" direction because of this uncertainty.
- For GT incidents, one value (2012) was slightly below the baseline and all others were above. This was judged to be a trend in the "wrong" direction, even more than simply continuing the upward trend.

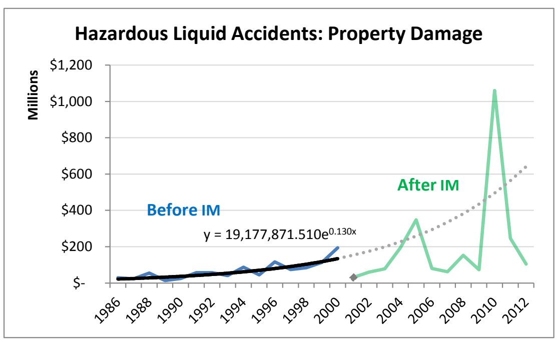
All other cases showed annual variation on both sides of the baseline after IM implementation; these were judged to be "inconclusive." In some of these cases, particularly deaths and major injuries, the numbers are simply too small to detect the kinds of patterns we were looking for. In other cases, the data might *suggest* either a positive or negative trend, but relaxing the evaluation criteria does not significantly affect the overall patterns. In any case, this is simply *an initial evaluation* of the trends to help guide the evaluation of external factors, program design, and program implementation.

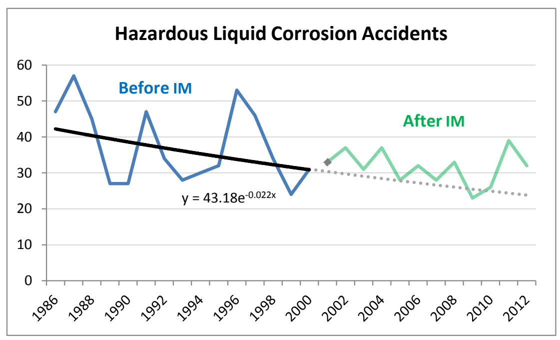

Hazardous Liquid Pipeline Accidents
Before IM
After IM
Pre-IM Trendline
forecast thru 2012
y = 182.59e^-0.016x

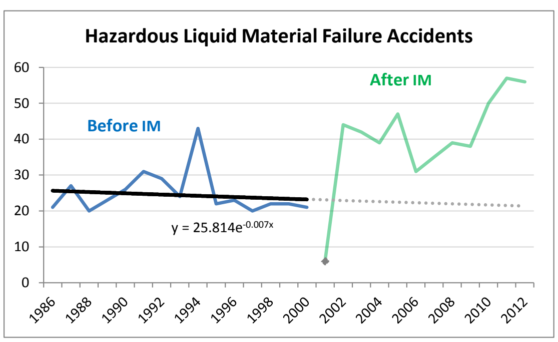

Hazardous Liquid Pipeline HC Accidents
Before IM
y = 32.309e^0.006x
After IM

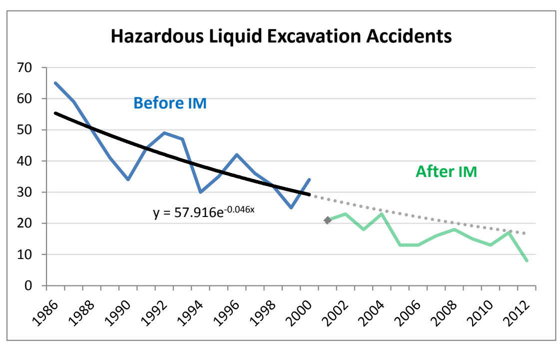

Hazardous Liquid Pipeline Deaths
Before IM
y = -0.068x + 2.943
After IM

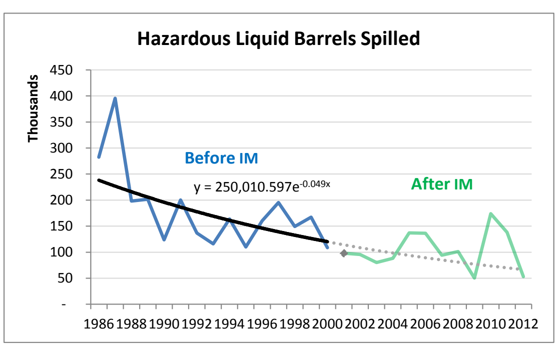

Hazardous Liquid Pipeline Injuries
Before IM
y = 27.589e^-0.100x
After IM

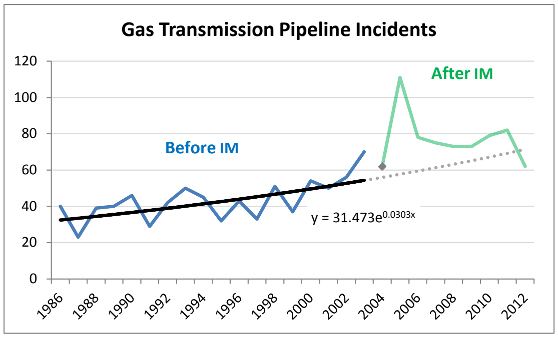

Hazardous Liquid Accidents: Property Damage
Millions
Before IM
y = 19,177,871.510e^0.130x
After IM

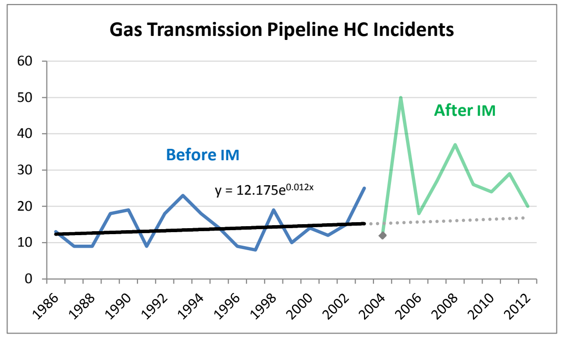

Hazardous Liquid Corrosion Accidents
Before IM
After IM
y = 43.18e^-0.022x

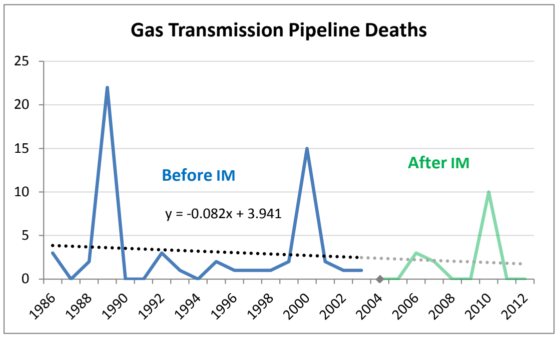

Hazardous Liquid Material Failure Accidents
After IM
Before IM
y = 25.814e^-0.007x

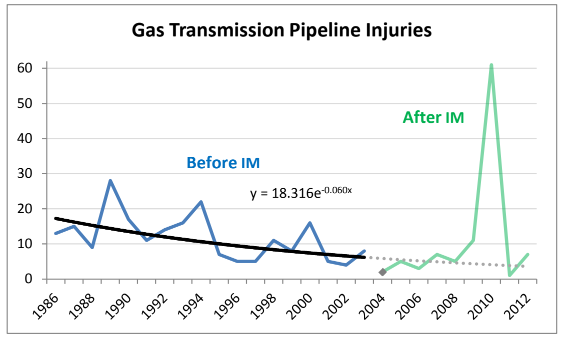

Hazardous Liquid Excavation Accidents
Before IM
After IM
y = 57.916e^-0.046x

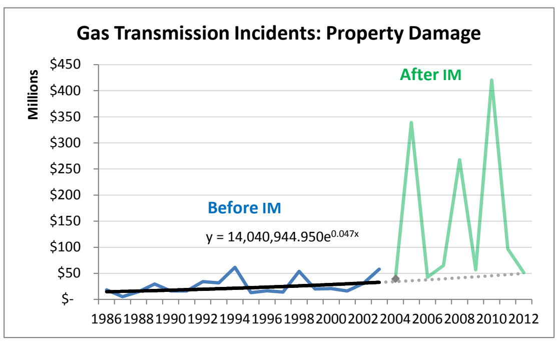

Hazardous Liquid Barrels Spilled
Thousands
Before IM
y = 250,010.597e^-0.049x
After IM

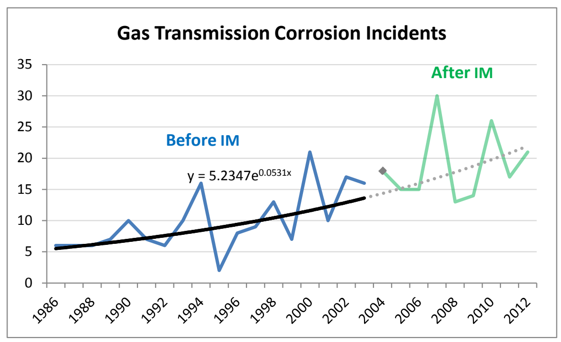

Gas Transmission Pipeline Incidents
After IM
Before IM
y = 31.473e^0.0303x

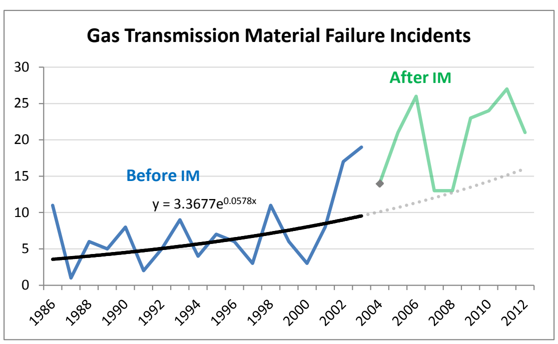

Gas Transmission Pipeline HC Incidents
After IM
Before IM
y = 12.175e^0.012x

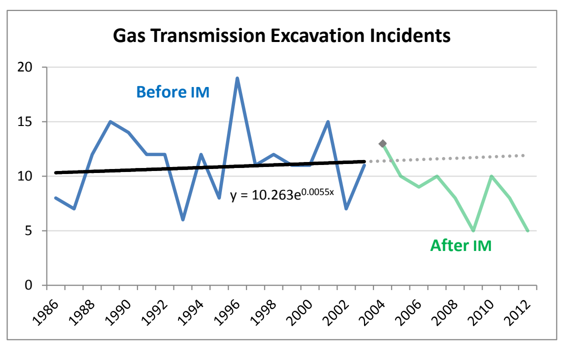

Gas Transmission Pipeline Deaths
Before IM
After IM
y = -0.082x + 3.941


Gas Transmission Pipeline Injuries
After IM
Before IM
y = 18.316e^-0.060x


Gas Transmission Incidents: Property Damage
Millions
After IM
Before IM
y = 14,040,944.950e^0.047x


Gas Transmission Corrosion Incidents
After IM
Before IM
y = 5.2347e^0.0531x


Gas Transmission Material Failure Incidents
After IM
Before IM
y = 3.3677e^0.0578x


Gas Transmission Excavation Incidents
Before IM
y = 10.263e^0.0055x
After IM

Limitations of the data and analysis

Most analyses confront issues with the data that can introduce errors into the findings or otherwise limit the conclusions that can be drawn. Here are some of the data issues that limited this evaluation, with a discussion of how these issues were handled. The aim is to provide enough information so that other analysts can replicate the findings, understand the sources of possible error, and (hopefully) work to improve the data with use.

There are several, general kinds of issues:

- The quality and completeness of the data reported by companies
- Comparability, and changes in reporting, over time
- Interpretation of the data

Quality and completeness

Incident data are collected and reported by pipeline operators. The data are reviewed by data analysts and inspectors for internal consistency and any obvious errors. But PHMSA investigates only a few dozen of these cases each year. In most cases, the agency relies on the operator to provide reliable data and reasonable judgments or conclusions about the causes, circumstances, and impacts that were relevant.

Cause data, in particular, involve a significant amount of professional judgment and when multiple causal factors are present, the choice of a primary cause might be subjective and other causal elements may be lost. This can have the effect of underestimating the prevalence of causes that often accompany other causes. So all of the breakouts by cause should be interpreted with some caution.

- In 2002, PHMSA revised the cause categories for incidents, which resulted in many fewer incidents reported with "other/unknown" causes. These were probably not evenly distributed across all cause categories. The data suggest that many of the incidents previously attributed to other causes would now be attributed to material failure.

  This was an important consideration in evaluating the HL incidents by cause, as it appeared that the increase in material failure after IM implementation (also beginning in 2002) might be explained by this change in reporting. For GT incidents, the increase in material failure incidents was larger than the decline in "Other" causes, so it was not judged to be a sufficient explanation.

HCA data might not fully reflect the number of incidents that occurred in, or affected, HCAs. In 2002, PHMSA began asking operators to identify whether the release was in an HCA. For spills above 5 barrels, this was a required field; below 5 barrels it was not. Several analysts have pointed out that the data field and reporting guidance were ambiguous on how to report this. Does it mean the release was from a pipeline segment in an HCA? From a segment that "could affect" an HCA? Or that it actually reached an HCA from a segment that "could affect" an HCA? This was clarified in 2010 guidance and reporting changes, but it does not appear to have had a large effect on reporting. This evaluation used a simplifying assumption that individual operators were more or less consistently reporting HCA data, so while comparisons across companies might be affected, comparisons overall across time should be less so.

Comparability over time

This evaluation used incident data collected from 1986-2012.

Incident reporting criteria were revised in the early 1980s for both HL and GT systems. Reporting criteria were further modified for HL systems in 1991, 1994, 1996, and 2002; and for GT systems in 2011. Some of these changes were significant in terms of comparability of the data over time.

- In 1986, incident reports were required for

Natural gas pipelines:	Hazardous liquid pipelines:
an event that involves a release of gas from a pipeline or … from an LNG facility and • a death, • personal injury necessitating in-patient hospitalization, or • estimated property damage [including cost of gas lost] of the operator or others or both of $50,000 or more; an event that results in an emergency shutdown of an LNG facility; or an event that is significant, in the judgment of the operator, even though it did not meet [other] criteria.	Any failure of a pipeline systems in which there is a release resulting in • death, • bodily harm resulting in loss of consciousness, necessity to carry the person from the scene, necessity for medical treatment, or disability which prevents the discharge of normal duties or the pursuit of normal activities beyond the day of the accident, • estimated property damage to the property of the operator or others or both exceeding $5,000, • explosion or fire not intentionally set by the operator, • loss of 50 barrels or more of liquid, or • escape to the atmosphere of more than 5 barrels a day of highly volatile liquids.

- In 1991, releases of carbon dioxide (CO_2) were added to the HL reporting criteria.

- In 1994, the cost of cleanup and recovery, and the value of lost product, were added explicitly to the total cost of property damage for HL accidents, and the reporting threshold for property damage was raised from $5,000 to $50,000—consistent with the criteria for natural gas systems.

By normalizing the data to $50,000 in property damage (in 1986 dollars), most of the effect from these changes was factored out.

- In 2002, the reporting threshold for HL spills was reduced from 50 barrels (2,100 gallons) to 5 gallons (with limited exceptions for maintenance). The definition of a reportable injury was simplified (and the threshold was raised) to any injury requiring in-patient hospitalization—consistent with the criteria for natural gas systems.

  The change in the volume criterion had a modest effect on (increasing) the total amount spilled, but a much more significant effect on the number of incidents reported beginning in 2002. However, for this analysis, I used only the higher reporting threshold that was common to both periods (before/after 2002). This is part of the data normalization used in constructing "significant" incidents.

  The change in injury reporting should have reduced, to some extent, the reporting of injuries from pipeline accidents or incidents.

- In 2011, for GT systems, property damage was redefined to exclude the cost of gas lost, and a new criterion was added: any unintentional release of gas with an estimated loss of three million cubic feet or more.

  This change had the effect of breaking the time series, as there is no easy way to compare the aggregated data before 2010 with the disaggregated data and new reporting criteria after 2010. Incidents with death, injury, or large property damages were not affected.

Dollar values in the incident data have been adjusted for overall inflation, but this does not take account of different rates of inflation for energy products or real estate values. I did a separate analysis of the change in natural gas prices alongside the changes in the number of reported incidents, and found some evidence that the increase in reporting might be largely attributed to these price changes. But the prices fluctuated widely, and without detailed data on each incident, there is no way to check the prices on the day any given incident occurred.

Data interpretation

Maybe the most important caution is that statistical patterns showing correlations or relationships in the data *do not prove causation*. None of the trends are proof that IM is or isn't working. They are simply indicators, suggesting a deeper exploration and analysis of the issues.

There is a possible time lag in the effects of the IM program, particularly as the program was implemented over a period of about seven years for one full cycle of assessments. This is not obvious from a review of the data/trends since IM implementation. But the general expectation is that improvements would be gained over time, not immediately in a step decline on the effective date of the regulations.

Trends in the data are typically represented using exponential regression to help reflect the logical concept of diminishing returns over time. Trend analysis is generally preferred to comparing averages before/after an intervention (like IM implementation) as it takes account the most recent data in establishing a baseline for comparison.

The period of transition from pre-IM to post-IM was not as exact as the one-year transition period assumed for the data analysis. Using a single, full calendar year was a simplifying assumption.

Small numbers—particularly for deaths and injuries—limit our ability to detect real trends. The "noise" (or normal fluctuations in the data) can exceed the "signal" we are looking for. This might be the most important reason why we don't see any patterns in the death and injury data.

Normalizing for changes in mileage or other external factors is a common way to help explain patterns in time series data. This was evaluated separately in analyzing some possible explanations for the safety outcomes we are seeing.

Disaggregating the data can help identify risk concentrations, particularly where a program like IM might be focused unequally on certain kinds of risk. Some people have suggested that breaking out line pipe failures from tank failures or facility-related incidents, or breaking out incidents on the right-of-way from those on operator property, might be more reflective of the efforts under the IM program. The data are not easy to analyze along these lines. And the IM program scope was not limited to these kinds of failures. Nevertheless, these kinds of analysis might be useful in understanding what has happened.

Other data issues

The repair rates—particularly immediate repairs per 1,000 miles of pipeline assessed—cannot be determined reliably from the data we have, although this is an obvious indicator of whether operators targeted their highest risk pipe for assessment first.

Reassessment intervals, particularly for liquid pipelines, might be too short. If these intervals are too conservative, the safety we are getting could be at a cost that is too high. But analyzing this would require data beyond what we have, so it beyond the scope of this evaluation. Further research might look at the actual growth of defects between successive assessments, across the industry.

If operators were targeting the highest-risk segments of their pipeline systems for assessment, we should have seen *a decline in the rate* of defects even over the baseline assessment period. Unfortunately, we didn’t collect the right data to tell us whether this was happening or not.

Annual reports don’t identify the number of anomalies *discovered* during the CY – only those repaired. While most repairs might be done in the same CY, there is some lag, and this makes it impossible to connect anomalies discovered with assessments (as in anomalies detected per 1,000 miles of pipeline)—normalizing the data. This is the measure that should be tracked.

Baseline miles completed (for HL) through 2010 total 114,896 miles (46% more than the total HCA miles of 78,500 in 2010) – suggests mileage was not reported correctly.

PHMSA didn’t start collecting HL assessment mileage data until 2004, which grouped together assessments conducted in 2002, 2003, 2004, and up to five prior years. Repairs were only for 2004 (evidently) so repairs per mile can’t be calculated until 2005, after 91% of the mileage was reportedly assessed for the baseline.

PHMSA stopped collecting MAOP and incident pressure for reportable incidents in 2010. This makes it nearly impossible (with the way data are currently collected) to evaluate the safety margins in practice.

Appendix C – Findings from other reviews and audits

Inspector General Audit (2012): Hazardous Liquid Pipeline Operators' Integrity Management Programs Need More Rigorous PHMSA Oversight

GAO reported that PHMSA had accumulated a backlog of IM inspections, and had not performed sufficient onsite visits to hazardous liquid pipelines and facilities. The audit also found that PHMSA's oversight of non-line pipe facilities were limited by less rigorous IM requirements even though these facilities account for more than half of all hazardous liquid accidents. GAO noted that PHMSA had not resolved longstanding data management deficiencies, had not established meaningful analysis capabilities, and had not established performance measures for assessing the program's effectiveness. Finally, GAO reported that the agency lacked the capability to identify high risk pipelines by linking accidents, inspection histories, and pipeline characteristics to their geographic location.

TRB Report (2012): Evaluating the Effectiveness of Offshore Safety and Environmental Management Systems

A TRB Panel reviewed the Minerals Management Service's inspection program for offshore facilities to assess its effectiveness in protecting human safety and the environment. The study included a review of many existing regulatory approaches used in the U.S. and internationally that are collectively referred to as safety management systems (SMS). The panel concluded that the goal should be a culture of safety in the industry, and that BSEE needs an evaluation system that focuses on attitudes and actions rather than documentation. It argued that companies are responsible for conducting SMS audits, but that the routine presence of competent inspectors was essential to verify compliance and assess safety culture. It pointed out some key differences in the skills needed for inspection vs. auditing in an SMS environment.

Congressional Research Service Report (2011): Keeping America's Pipelines Safe and Secure: Key Issues for Congress

CRS evaluated several issues that might be relevant for program reauthorization. In assessing the need for higher civil penalties, it noted that penalties, even if raised substantially, would account for only a limited share of the overall financial impact from a pipeline release. By contrast, PHMSA's ability to influence pipeline operations through corrective action orders or shutdown orders has been much more substantial, and these can have large financial impacts. The CRS analysis also reinforced the issue of data quality, as noted by NTSB in its investigation of the San Bruno incident; and discussed some of the technical issues and tradeoffs in requiring more internal assessment ("pigging") of pipelines. Finally, it discussed some of the background on the issue of allowing pipeline integrity reassessment intervals to be changed from fixed seven-year intervals to intervals based on technical data, risk factors, and engineering analysis, as recommended by GAO in 2006.

Inspector General Review (2006): Integrity Threats to Hazardous Liquid Pipelines

In 2006, the Office of the Inspector General reviewed actions taken by hazardous liquid pipeline operators to remediate integrity threats and OPS efforts to verify the adequacy of these corrective actions. The OIG found that repairs were completed as reported, 98% of repairs were completed within established timeframes, and OPS and/or states had inspected the integrity management programs of 86% of the operators covering 98% of all pipeline mileage in or potentially affecting HCAs. The review also found inaccuracies in operators' annual reports, and errors in operators' analysis and interpretation of smart pig results. OPS had already taken actions to respond to these challenges when the report was issued.

GAO Audit (2006): Risk-Based Standards Should Allow Operators to Better Tailor Reassessments to Pipeline Threats

GAO found that reassessments of gas transmission pipelines were useful, but that the 7-year reassessment interval (for corrosion) appeared to be conservative. GAO recommended that Congress consider allowing gas transmission operators to reassess their pipelines using risk-based standards instead.

GAO Audit (2006): Integrity Management Benefits Public Safety, but Consistency of Performance Measures Should be Improved

GAO evaluated the IM program's effects on public safety, and the agency's (and states') plans to oversee operator implementation of program requirements. GAO found some early indications that the condition of pipelines was improving, based on mileage assessed and the number of repairs completed, and that operators were making good progress in assessing pipelines and making repairs, but that they needed to better document their decisions and processes. GAO reported that the usefulness of performance measure data was limited by inconsistencies in reporting causes of failures, and recommended that PHMSA revise its incident reporting requirements to account for changes in the price of natural gas.

GAO Audit (2002):

GAO examined PHMSA's approach to IM and found that PHMSA had experienced a number of problems with data completeness and accuracy. GAO also reported that PHMSA accident reporting contained too few causal categories. As a result, PHMSA implemented several changes to enhance data quality, and increased the number of causal categories from seven to twenty five.

Appendix D - Bibliography

Legislation

1. *Accountable Pipeline Safety and Partnership Act of 1996* (Public Law 104-304, October 12, 1996)
2. *The Pipeline Safety Improvement Act of 2002* (P.L. 107-355)
3. *The Pipeline Integrity, Protection, Enforcement and Safety Act of 2006* (P.L. 109-468)
4. *Pipeline Safety, Regulatory Certainty, and Job Creation Act of 2011* (January 3, 2012)
5. *49 USC 60109 – High Density Population Areas and Environmentally Sensitive Areas*

Federal Regulations and Rulemaking

6. *49 CFR Part 195 Pipeline Safety: Pipeline Integrity Management in High Consequence Areas (Hazardous Liquid Operators with 500 or More Miles of Pipeline), Proposed Rule* (Federal Register April 24, 2000)
7. *49 CFR Part 195 Pipeline Safety: Pipeline Integrity Management in High Consequence Areas (Hazardous Liquid Operators with 500 or More Miles of Pipeline), Final Rule* (Federal Register December 1, 2000)
8. *US DOT/RSPA Final Regulatory Evaluation* – Pipeline Integrity Management in High Consequence Areas, Docket RSPA-99-6355-23
9. *49 CFR Part 192 – Pipeline Safety: High Consequence Areas for Gas Transmission Pipelines; Notice of Proposed Rulemaking* (Federal Register, January 9, 2002)
10. *49 CFR Part 192 – Pipeline Safety: High Consequence Areas for Gas Transmission Pipelines; Final Rule* (Federal Register, August 6, 2002)
11. *49 CFR Part 192 - Pipeline Safety: Pipeline Integrity Management in High Consequence Areas (Gas Transmission Pipelines); Proposed Rule* (Research and Special Programs Administration, January 28, 2003)
12. *49 CFR Part 192 - Pipeline Safety: Pipeline Integrity Management in High Consequence Areas (Gas Transmission Pipelines); Final Rule* (Research and Special Programs Administration, December 15, 2003)

13. *US DOT/RSPA Final Regulatory Evaluation* – Pipeline Integrity Management in High Consequence Areas (Gas Transmission Pipelines), Docket RSPA-00-7666-356

14. *29 CFR 1910.119 – Process Safety Management of Highly Hazardous Chemicals* (Occupational Safety and Health Administration)

15. *Advance Notice of Proposed Rulemaking: Safety of On-Shore Hazardous Liquid Pipelines,* 76 FR 303 published October 10, 2010

16. *Advance Notice of Proposed Rulemaking: Safety of Gas Transmission Pipelines,* 76 FR 5308 published August 25, 2011

Standards

17. *API Standard 1160 – Managing System Integrity for Hazardous Liquid Pipelines (2001)*

18. *The Department of Transportation's Information Dissemination Quality Guidelines (2002) - http://docketsinfo.dot.gov/ombfinal092502.pdf*

19. *ASME B31.8S Managing System Integrity of Gas Pipelines (2004 and 2010 editions)* – American Society of Mechanical Engineers

20. *A Guide to the Offshore Installations (Safety Case) Regulations 2005 (2006)* – Health and Safety Executive (UK)

21. *Guidance on Risk Assessment for Offshore Installations: Offshore Information Sheet No. 3/2006 (2006)* – Health and Safety Executive (UK)

22. *ISO 31000:2009 Risk Management – Principles and Guidelines*, International Organization for Standardization

23. *ISO/IEC 31010:2009 Risk Management – Risk Assessment Techniques*, International Organization for Standardization

24. *Executive Order 13563: Improving Regulation and Regulatory Review,* President Barack Obama, January 18, 2011, published in the Federal Register January 21, 2011

25. *International Safety Management (ISM) Code (2010)* – International Maritime Organization

26. *DOT Safety Management Systems Guidance Document* (2012) – Secretary of Transportation memo to Modal Administrators

Program Documentation

27. *Gas Integrity Management Protocol Diagrams* (PHMSA Website, January 18, 2006)

28. *Integrity Management Inspection Protocols [for Hazardous Liquid Pipelines]*(December 2007)

29. *Gas Integrity Management Inspection Manual: Inspection Protocols with Results Forms* (January 1, 2008)

30. *Enforcement Guidance for Liquid and Gas Transmission IM* (October 2008)

31. *Hazardous Liquid Integrity Management Progress Report* (DOT/PHMSA, January 2011)

32. *Gas Transmission Integrity Management Progress Report* (DOT/PHMSA, February 2011)

33. *U.S. Department of Transportation Call to Action to Improve the Safety of the Nation's Energy Pipeline System* (Revised November 2011)

34. *Oversight of Performance-Based Regulations* (Office of Pipeline Safety presentation to the DOT Safety Council, February 24, 2012)

35. *Frequently Asked Questions for Integrity Management*, posted on the PHMSA website

Program Reviews

36. *Oil Pipeline Characteristics and Risk Factors: Illustrations from the Decade of Construction* (Kiefner and Trench for American Petroleum Institute, December 2001)

37. *GAO Audit (2002):*

38. *The U.S. Oil Pipeline Industry's Safety Performance* (Cheryl J. Trench for the American Association of Oil Pipelines, February 2003)

39. *Integrity Threats to Hazardous Liquid Pipelines* (Inspector General Report AV-2006-071, September 18, 2006)

40. *Natural Gas Pipeline Safety: Risk-Based Standards Should Allow Operators to Better Tailor Reassessments to Pipeline Threats* (GAO 06-945, September 2006)

41. *Natural Gas Pipeline Safety: Integrity Management Benefits Public Safety, but Consistency of Performance Measures Should Be Improved* (GAO 06-946, September 2006)

42. *Investigate Fundamentals and Performance Improvements Of Current In-Line Inspection Technologies For Mechanical Damage Detection,* prepared for PHMSA and Pipeline Research Council International, Inc. by Blade Energy Partners, May 2008 and July 2009

43. *Keeping America's Pipelines Safe and Secure: Key Issues for Congress* (Paul Parfomak, Congressional Research Service Report for Congress, March 17, 2011)

44. *Hazardous Liquid Pipeline Operators' Integrity Management Programs Need More Rigorous PHMSA Oversight* (Office of Inspector General Report AV-2012-140, June 18, 2012)

45. *Evaluating the Effectiveness of Offshore Safety and Environmental Management Systems* (Transportation Research Board Special Report 309, 2012)

46. *Track Record of In-Line inspection as a Means of ERW Seam integrity Assessment (Kiefner Associates and Det Norske Veritas report for PHMSA, November 15, 2012)*

Risk Literature

47. *Judgment Under Uncertainty: Heuristics and Biases* – Amos Tversky and Daniel Kahneman (1982)

48. *Managing the Risks of Organizational Accidents* – James Reason (1997).

49. *Understanding Risk: Informing Decisions in a Democratic Society* – National Research Council, Committee on Risk Characterization (1996).

50. *Normal Accidents: Living with High-Risk Technologies* – Charles Perrow (1999).

51. *The Regulatory Craft: Controlling Risks, Solving Problems, and Managing Compliance* – Malcolm K. Sparrow (2000).

52. *Barriers and Accident Prevention* – Erik Hollnagel (2004).

53. *Pipeline Safety Risk Management: Ideas, techniques, and Resources* (W. Kent Muhlbauer, Third Edition, published by Elsevier, 2004)

54. *Managing the Unexpected* – Karl E. Weick and Kathleen M. Sutcliffe (2007).

55. *Updated Principles for Risk Analysis* (OMB Memorandum M-07-24, September 19, 2007)

56. *What's Wrong with Risk Matrices?* (Louis Anthony Cox, Jr., *Risk Analysis*, Vol. 28, No. 2, 2008)

57. *Data-Driven Risk Models Could Help Target Pipeline Safety Inspections* – Bureau of Transportation Statistics Special Report SR-010 (Kowalewski and Young, July 2008).

58. *The Failure of Risk Management: Why It's Broken and How to Fix It* (Douglas W. Hubbard, published by John Wiley & Sons, Inc., 2009)

59. *Complex Systems and Human Behavior* (Christopher G. Hudson, published by Lyceum Books, Inc., 2010)

60. *Safety Culture:* A Research Paper prepared for the U.S. DOT Safety Council (May 2011).

61. *Progress on Process Safety Indicators – Necessary but not Sufficient? A Discussion Paper for the U.S. Chemical Safety and Hazard Investigation Board* (Peter Wilkinson, Noetic Risk Solutions Pty Ltd, 2012)

Accident Investigation Reports

62. *Pipeline Rupture and Subsequent Fire in Bellingham, Washington June 10, 1999* (National Transportation Safety Board, Accident Report NTSB/PAR-02/02)

63. *Natural Gas Pipeline Rupture and Fire near Carlsbad, New Mexico August 19, 2000* (National Transportation Safety Board, Accident Report NTSB/PAR-03/01)

64. *Pacific Gas and Electric Company Natural Gas Transmission Pipeline Rupture and Fire, San Bruno, California September 9, 2010* (National Transportation Safety Board, Accident Report NTSB/PAR-11/01)

65. *Enbridge Incorporated Hazardous Liquid Pipeline Rupture and Release, Marshall, Michigan July 25, 2010* (National Transportation Safety Board Accident Report, NTSB/PAR-12/01)

Endnotes

Executive summary

[1] Pacific Gas and Electric Company Natural Gas Transmission Pipeline Rupture and Fire, San Bruno, California September 9, 2010 (National Transportation Safety Board, Accident Report NTSB/PAR-11/01)

[2] 49 CFR Part 195 Pipeline Safety: Pipeline Integrity Management in High Consequence Areas (Hazardous Liquid Operators with 500 or More Miles of Pipeline), Final Rule (Federal Register December 1, 2000). This rule was effective March 31, 2001; operators were required to develop written program plans within one year of that date. For this evaluation, 2001 was considered a transition year.

[3] 49 CFR Part 192 - Pipeline Safety: Pipeline Integrity Management in High Consequence Areas (Gas Transmission Pipelines); Final Rule (Research and Special Programs Administration, December 15, 2003). This rule was effective January 14, 2004; operators were required to have written program plans by Dec. 17, 2004. For this evaluation, 2004 was considered a transition year.

[4] DOT Strategic Plan 2006-2011

[5] This includes field inspectors, region directors, program managers, the pipeline industry, and outside stakeholders. At the outset of each interview, they were asked: Do you believe the IM program is effective? Nearly everyone responded affirmatively, most with similar reasoning.

[6] Intentionally not defining the term when the question was posed – to begin the conversation about why they believed it was, or wasn't, effective.

[7] Almost two-thirds of the reduction in the total number of hazardous liquid accidents can be attributed to a decline in excavation damage – see discussion on page 37.

[8] This issue has been identified by the program already (before this evaluation) and might be addressed in a current rulemaking. But some inspectors expressed serious concerns about the potential consequences of leaving these standards in place for any length of time.

[9] See the Advance Notices of Proposed Rulemaking for both hazardous liquid and gas transmission pipelines. These are discussed in more detail in the section on Program Management.

Context – *What shaped the program*

[10] From operators' annual reports to PHMSA.

[11] The terms are used interchangeably in this report, although traditionally (and in regulation) "accidents" refer to failures of liquid pipelines and "incidents" refer to failures of natural gas pipelines.

[12] States generally oversee all intrastate gas pipelines (110,00 miles), and ten states inspect interstates gas transmission systems (about 40,000 miles) under agreement with PHMSA. Five of these states also inspect hazardous liquid pipelines (about 14,000 miles) under agreement with PHMSA.

[13] Including accidents in Edison, NJ (1994) and Fork Shoals, SC (1996)

[14] Pipeline Rupture and Subsequent Fire in Bellingham, Washington June 10, 1999 (National Transportation Safety Board, Accident Report NTSB/PAR-02/02)

[15] In fact, the Notice of Proposed Rulemaking was published before API-1160 was completed, so it was not adopted. The final version of API-1160 included repair criteria that were less stringent than both draft versions and those included in the final rule.

[16] Natural Gas Pipeline Rupture and Fire near Carlsbad, New Mexico August 19, 2000 (National Transportation Safety Board, Accident Report NTSB/PAR-03/01)

[17] Jeff Wiese video on DIMP Introduction, from PHMSA website.

Program logic – *how the program works*

[18] Based on analysis of the rulemaking documents, program documentation, and discussion with program managers and inspectors.

[19] *Ibid.*

[20] Drawn from Budget Requests (FY 2002-2013) to the Committees on Appropriations.

[21] Based on analysis of the rulemaking documents, program documentation, and discussion with program managers and inspectors.

Program outcomes – *the expected results*

[22] For gas transmission systems, class location might be considered a proxy for population HCAs. But there are significant data limitations in using this for analysis. Relatively little GT mileage is classified as in HCAs, so there are few HCA incidents reported even after IM implementation. Class location has not been reported for every incident, and in any case it is not directly correlated with HCAs – about 90% of incidents in class 4 locations were also in HCAs, and about 50% of incidents in class 3 locations were; a small fraction of class 2 and class 1 incidents were as well.

The data – *what actually happened*

[23] Green check marks (✓) indicate trends going in the "right" direction with respect to expectations; red "X" marks (×) indicate trends going in the "wrong" direction. See Appendix B for details on the methodology for normalizing the data over time and determining *meaningful* trends.

[24] This could be a simple artifact of changes in reporting over time. In the early years of IM implementation, the guidance for liquid pipeline operators did not clearly differentiate between accidents in an HCA vs. accidents in pipeline segments that could affect an HCA, but were otherwise outside the HCA boundary. Of course, this also highlights an issue with the reliability of the data collected and used to manage the program.

[25] Any incident involving a death, major injury, property damage of $500,000 or more (in 2012 dollars), or a liquid release of 100,000 gallons or more (a "major" spill) – see Appendix B, Methodology and Limitations, for more details.

[26] Two of these (in August 2005) were related to Hurricane Katrina, and one (in February 2008) was related to a tornado in Hartsville, TN. Incidents caused by natural force damage are sometimes viewed as less "controllable," but they are reportable incidents under the regulations and operators are expected to take natural force damage into account in their risk assessments.

[27] Defined here as any incident involving a death, major injury, property damage of $500,000 or more (in 2012 dollars), or a liquid release of 100,000 gallons or more (a "major" spill as defined in the Federal On Scene Coordinator's Guide to Environmental Response). These latter two criteria are each about 10% of the total number of reported incidents, with significant overlap.

[28] Corrosion and material/weld failure are generally considered "IM-detectable" failures since these are targeted particularly in the required assessments and with in-line inspection tools under integrity management.

[29] Less than one percent increase per year (2002-2011) for hazardous liquid pipelines, and less than one percent total for gas transmission pipelines over the same 10-year period.

[30] Largely due to difficulties in estimating, in advance, which pipeline segments "could affect" a high consequence area. These determinations required more detailed information about geography and hydrology along with geospatial data on several kinds of high consequence areas.

[31] This is consistent with the findings from Kiefner and Trench (2001) and Trench (2003), except that it provides an additional estimate of the failure rates from 0-6 years old, at which point failures reach the longer-term "base rate."

[32] http://www.census.gov/construction/nrc/historical_data/

[33] http://www.census.gov/geo/www/ua/2010urbanruralclass.html

[34] There is a substantial body of literature on risk perception, including the observation that people "overestimate" risks that are unfamiliar, beyond their control, or especially gruesome.

[35] For significant accidents, the trendline through 2001 shows a level of 140/year total as the baseline. The average number of significant accidents per year from 2002-2011 was 122—an average reduction of 18/year. Similar calculation for excavation damage incidents indicates an average reduction of 9/year.

Program design

[36] Pipeline Safety: Pipeline Integrity Management in High Consequence Areas (Hazardous Liquid Operators with 500 or More Miles of Pipeline), Final Rule, Federal Register December 1, 2000.

[37] Executive Order 13563: Improving Regulation and Regulatory Review

[38] "Any measure that shifts the onus for maintaining safe work practices on to the organizations directly concerned has to represent an enormous 'plus' in the struggle to limit organizational accidents." – James Reason, in Managing the Risks of Organizational Accidents, (p.181)

[39] From an Overview and Comparison of Quality Management Systems by Jim von Hermann, October 2012. Note, too, that NTSB added safety management systems as one of its Most Wanted safety improvements in 2012.

[40] There is at least one statutory exception to this -- The Federal Water Pollution Control Act of 1972, as amended prohibits the discharge of any harmful quantity of oil or hazardous substances into U.S. waters. Under the law, any spill into U.S. waters resulting in a visible sheen is subject to a civil penalty. The law provides for pollution prevention regulations as well, any many regulations have been issued, but the basic prohibition and penalty provisions for a discharge set one of the most outcome-oriented standards in place anywhere. The U.S. Coast Guard and EPA have enforcement responsibility for coastal and inland waters, respectively.

[41] See TRB Special Report 309 – *Evaluating the Effectiveness of Offshore Safety and Environmental Management Systems*.

[42] Secretary of Transportation memorandum to Modal Administrators (undated) – DOT Safety Management Systems Guidance Document.

[43] Texas City meeting with oil and gas industry; TRB report: *Evaluating the Effectiveness of Offshore Safety and Environmental Management Systems*

[44] See *Evaluating the Effectiveness of Offshore Safety and Environmental Management Systems* (Transportation Research Board Special Report 309, 2012); also based on interviews with field inspectors, risk modeling practitioners, and comments at the Texas City meeting with the oil and gas industry.

[45] From interviews with field inspectors and many comments at the Texas City meeting with the oil and gas industry.

[46] For liquid pipelines, test pressure is 1.25 time MOP for 4 hours with an additional 1.1 times MOP for 4 hours if the pipe cannot be observed. For gas transmission pipelines, test pressure varies by class location—ranging from 1.1 times MAOP for class 1 locations to 1.5 times MAOP for class 3 and 4 locations.

[47] Or—for hazardous liquid systems—maximum operating pressure (MOP), with essentially the same meaning as MAOP. In this report, the term MAOP is used for both, simply for convenience.

[48] From observations by many pipeline inspectors and region directors. Also see Muhlbauer, page 106.

[49] See PHMSA/PRCI report: *Investigate Fundamentals and Performance Improvements Of Current In-Line Inspection Technologies For Mechanical Damage Detection* (Phase 1 report), prepared by Blade Energy Partners, May 2008.

[50] FAQ 7.19: Information on tool tolerances should be used to assure that defects requiring early excavation and mitigative action are properly identified and characterized. This does not necessarily mean simply adding the vendor-supplied tolerance value to reported depth of indications. Several sources of data may be used …

[51] A review of burst test data by Pipeline Research Council International, Inc. (PRCI) in 2000 (PRCI Report Catalog Number L51878) raised concerns that use of these methods can, in some instances, result in predicted failure pressures that are greater than the recorded burst pressures from actual tests.

[52] See Executive Summary in *Track Record of In-Line inspection as a Means of ERW Seam integrity Assessment* (Kiefner Associates and Det Norske Veritas report for PHMSA, November 15, 2012) – the researchers concluded that “Among the 13 cases examined, there was no case for which the investigating team is willing to say that the inspection provided full confidence in the seam integrity of the assessed segment.”

[53] 49 CFR 192.933(d) does not specifically state this condition, but requires following ASME Standard B31.8S which does list 80% metal loss as an immediate repair condition.

[54] 49 CFR 195.452(h)(4) and FAQ 7.18

[55] Operators interpreted the reporting guidance in different ways, so we do not have good data on the amount of mileage assessed from 2002-2009. The data appear to show more baseline miles assessed than there were miles to assess.

Program implementation

[56] Based on interviews with inspectors, industry safety personnel, and former industry employees.

[57] Conclusions drawn from interviews with PHMSA inspectors and region directors.

[58] This is not to say that no PHMSA employees have any expertise in this, but it is not institutional. There are no positions requiring it, and no evidence that this kind of expertise was used in developing the program.

[59] From interviews with inspectors and program staff who were deeply involved in the program implementation,.

[60] Most of the observations here are supported by interviews with 10 inspectors and 5 region directors across the country—all intimately familiar with the IM program, and collectively drawing on about 100 years of experience inspecting and assessing companies’ IM programs. This includes some experience working in the regulated industry before coming to PHMSA.

[61] Based on interviews with inspectors, industry safety personnel, and former industry employees.

[62] Safety Culture: A Research Paper prepared for the U.S. DOT Safety Council (May 2011).

[63] James Reason, in *Managing the Risks of Organizational Accidents*.

[64] FAQ 8.18 (liquid pipelines) and FAQ-102 (gas pipelines).

[65] Explained in *Normal Accidents: Living with High-Risk Technologies*, by Charles Perrow.

[66] James Reason, in *Managing the Risks of Organizational Accidents* (pp.85-105).

[67] Ibid. (p.51)

[68] Ibid. (p.55)

Risk modeling and risk assessment

[69] 49 CFR 195.452(e)(1).

[70] FAQs elaborate on this – for example, in FAQ 1.5 (for liquid pipelines): Recognized industry practices include those found in national consensus standards or reference guides. In FAQ 1.6: Use of an alternative must provide an equivalent (or better) result than using the recognized practice.

[71] Pipeline Risk Management Manual: Ideas, Techniques, and Resources (Third Edition), by W. Kent Muhlbauer (2004)

[72] In fact, many standards provide these same general choices as acceptable ways of evaluating risk – including *ISO/IEC 31010:2009 Risk Management – Risk Assessment Techniques,* and *Guidance on Risk Assessment for Offshore Installations: Offshore Information Sheet No. 3/2006 (2006)* – Health and Safety Executive (UK).

[73] Based on interviews with PHMSA inspectors. Also noted in the ANPRM published August 25, 2011 for gas transmission pipelines.

[74] See *The Failure of Risk Management: Why It's Broken and How to Fix It* (Douglas W. Hubbard, published by John Wiley & Sons, Inc., 2009) for an extensive discussion of these weaknesses in index-scoring models.

[75] *Oil Pipeline Characteristics and Risk Factors: Illustrations from the Decade of Construction* (Kiefner and Trench for American Petroleum Institute, December 2001)

[76] *Data-Driven Risk Models Could Help Target Pipeline Safety Inspections* – Bureau of Transportation Statistics Special Report SR-010 (Kowalewski and Young, July 2008) for a discussion of the issues and a comparison of several approaches.

[77] Extensively demonstrated in research by Daniel Kahneman and Amos Tversky.

[78] *The Failure of Risk Management: Why It's Broken and How to Fix It.*

[79] See *What's Wrong with Risk Matrices?* (Louis Anthony Cox, Jr., *Risk Analysis*, Vol. 28, No. 2, 2008)

[80] *The Failure of Risk Management: Why It's Broken and How to Fix It.*

[81] ISO 31000, API 1160, ASME B31.8S, and Pipeline Risk Management Manual, among others.

[82] Adapted from *The Failure of Risk Management: Why It's Broken and How to Fix It.*

[83] Judgment and Uncertainty: Heuristics and Biases, by Amos Tversky and Daniel Kahneman (1982).

[84] See *Data-Driven Risk Models Could Help Target Pipeline Safety Inspections* – Bureau of Transportation Statistics Special Report SR-010 (Kowalewski and Young, July 2008) for a discussion of the issues and a comparison of several approaches.

[85] Ibid.

[86] OMB Memorandum M-07-24 – principles for risk assessment.

[87] See *Data-Driven Risk Models Could Help Target Pipeline Safety Inspections* – Bureau of Transportation Statistics Special Report SR-010 (Kowalewski and Young, July 2008) as a potential starting point.

Data quality

[88] This process is explored in much more detail in the Data Quality Assessment.

[89] FAQ 2.9 addresses certain kinds of information an operator should keep.

[90] FAQ 7.19 (liquid systems) and FAQ-68 (gas systems)

[91] FAQ 6.8

[92] FAQ-240

[93] Liquid Inspection Protocol 5.02, December 2007.

[94] Gas IM Protocols, Revision 5, 1/1/2008

[95] See the PHMSA Data Quality Assessment, 2009.

[96] The analysis of PIPP provides a good illustration of this – one variable (past accidents) was a relatively good predictor of risk. By combining it in a model with 10-11 weak risk factors, the effect was washing out the most useful (predictive) information the program had about a company's risk.

Metrics

[97] See *Managing the Risks of Organizational Accidents*, page 121.

[98] In comments on the proposed rule, NTSB commented that the requirement had to contain unequivocal guidance if operators are to use it to improve their programs, and suggested the agency develop measures. OPS simply responded that it had not revised the guidance.

[99] See Chemical Safety Board paper, and the DOT Information Quality Guidelines.

[100] See *Barriers and Accident Prevention*, by Erik Hollnagel, for a good discussion of barrier functions and use, and the role of barriers in accidents.

[101] See *Managing the Risks of Organizational Accidents*, page 126.

[102] Ibid, page 107.

[103] For example, to determine the likelihood of certain consequences given a failure, you need all failures. When failure data are screened based on the outcome you are looking for in the first place, it's impossible to determine the relative risks in the presence or absence of certain conditions.

[104] From *Managing the Risks of Organizational Accidents* (p.119)

[105] Ibid. (p. 114)

Inspection/oversight

[106] This risk model shared many of the same weaknesses as PIPP and RRIM, but is no longer used.

[107] Gas IM Protocols meeting Jan 2005, p. 64

[108] See FAQ 11.1

[109] FAQ 8.13

[110] From Inspection Program Briefing Sheet.

Enforcement

[111] The definition of safety culture, from the DOT Safety Council.

[112] There is a fairly extensive literature on enforcement and compliance, ranging across the disciplines of economics, law, and organizational psychology.

[113] See *Making Things Stick: Enforcement and Compliance*, by A. G. Heyes, Oxford Review of Economic Policy (1998)

[114] Most inspectors I interviewed.

[115] See, for example, *The Theory of Public Enforcement of Law*, by A Mitchell Polinsky and Steven Shavell, October 2005.

[116] This emphasis has been reflected at least as far back as 2006, and was expressed explicitly in the PHMSA Strategic Plan (2006-2011).

Program management

[117] See *Managing the Risks of Organizational Accidents*, by James Reason (p. 181)

[118] Managing the Risks of Organizational Accidents – James Reason (1997).

[119] Section 12(d) of the National Technology Transfer Act of 1995 (P.L. 104-113, or "the Act") codified the policies of Circular A-119. Section 12(d)(1) states that "Except as provided in paragraph (3) of this subsection, all Federal agencies and departments shall use technical standards that are developed or adopted by voluntary consensus standards bodies, using such technical standards as a means to carry out policy objectives or activities determined by the agencies and departments.

[120] See Federal Register December 1, 2000 (hazardous liquid IM rule), page 75382.

[121] Ibid, p. 75394 – In comments on the proposed rule, Fuel Safe Washington noted that "Appendix C [the risk model illustration] is completely undermined by allowing operators to apply their own weights or values to risk factors." OPS responded simply that it "continues to believe that the guidance in Appendix C will be helpful to operators …"

[122] See *Data-Driven Risk Models Could Help Target Pipeline Safety Inspections* – Bureau of Transportation Statistics Special Report SR-010 (Kowalewski and Young, July 2008).

[123] Framed by Jim von Hermann in a QMS overview discussion with the Associate Administrator for Pipeline Safety.

[124] See *Managing the Risks of Organizational Accidents*, by James Reason

[125] Explained in *Normal Accidents: Living with High-Risk Technologies*, by Charles Perrow.

[126] The DOT Safety Council completed an extensive literature review on safety culture in 2011. NTSB also has advocated for several years that transportation companies should develop safety management systems; SMS was on NTSB's "Most Wanted" list of safety improvements in 2012.

[127] This is a focus of some research being done now by the DOT Safety Council

[128] 76 FR 303 (for hazardous liquid pipelines) published October 10, 2010 and 76 FR 5308 (for gas transmission pipelines) published August 25, 2011