RISKOPILOT

RISKOPILOT BRIEFING · CRITICAL CONTROLS SERIES

Thirty clean trials prove less than you think

The pilot report says thirty trials, zero misses. It sounds like proof. The arithmetic says the system could still miss almost one time in ten.

BRUNO HOUNKPATI, CEO AND FORMER VP HSE

NOVEMBER 2026 · 6 MINUTES

A pilot report lands on the executive's desk. The proximity system was tested thirty times with a surrogate target. It detected the target thirty times. Zero misses. The recommendation is to fit the fleet.

Thirty out of thirty feels like certainty. It is a small sample, and small samples allow a good deal of bad news to hide.

The arithmetic

In 1983 James Hanley and Abby Lippman-Hand published a short paper in JAMA with a title worth remembering: "If nothing goes wrong, is everything all right?" Their answer became known as the rule of three. If an event has not occurred in n trials, the upper 95% limit on its true rate is about 3 divided by n.

Thirty trials, no misses: the true miss rate could be as high as about one in ten. The exact figure is 9.5%.

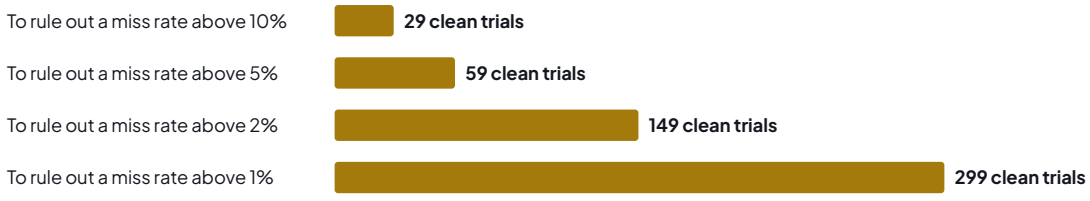
Turn the question round. How many clean trials does it take to rule out a given miss rate, at the same confidence?

- To rule out a miss rate above **10%**: **29** trials.
- Above **5%**: **59** trials.
- Above **2%**: **149** trials.
- Above **1%**: **299** trials.

These numbers are our calculation from the exact binomial formula. They assume every trial is independent and represents the conditions that matter.

© 2026 Riskopilot · riskopilot.com · contact@riskopilot.com · CC BY-ND 4.0

EXHIBIT



Smallest n for which 1 − 0.05^(1/n) falls at or below the target. Independent, representative trials, no misses, 95% one-sided.

The assumption that rarely holds

Independence is where most pilots fall short. The same mannequin, in the same spot, on the same clear afternoon, approached thirty times by the same driver, is closer to one trial than to thirty. The formula does not know that. It will give you 9.5% all the same.

So the honest reading of a typical thirty-run pilot is weaker than one in ten. It tells you the system can detect that target in those conditions. It tells you little about dusk, dust, the night shift or the subcontractor's machine.

Zero is a setting

There is a second reason to distrust a clean sheet. Any detector trades misses against false alarms. Set it wide and sensitive and it will miss less and alarm more.

The best field study we know of shows what that looks like. On one mine haul truck over seven days, a radar system alarmed on 580 of 618 reversing movements and missed nothing. Only 41% of those alarms called for action or caution. A pilot that reports misses and says nothing about alarms has reported half the result.

Ask for both denominators. Of all movements, how many produced an alert? Of all alerts, how many needed someone to act?

Late is not a pass

A third number is usually missing. A detection that arrives after the last moment at which the driver could have stopped is a miss for the person on foot, whatever the log says. Count alerts in time, alerts that were late, misses, and trials where the reference failed. That last group is unknown. It is never a pass.

The uncomfortable conclusion

Follow the arithmetic and you reach a result suppliers and buyers both dislike. The more you need to rely on a function, the less a site pilot can tell you.

If you want an advisory signal and can live with "probably misses less than one time in ten", thirty good trials will support you. If you want to rely on the function as the last signal before a fatal exposure, you need a miss rate far below 1%, and no site can run the hundreds of

© 2026 Riskopilot · riskopilot.com · contact@riskopilot.com · CC BY-ND 4.0

independent trials that claim requires.

That is the case for keeping a piloted function advisory. Relying on it as part of a control is an engineering decision, made on design evidence and a manufacturer's validation, with the site's technical authority. A pilot cannot make it for you.

The strongest case against this article

"This is paralysis by statistics. People are at risk tonight." A fair charge if the conclusion were "do nothing". It is the opposite. Fit the barrier, change the layout, and add the system as an advisory layer if you wish. What the arithmetic forbids is removing a control because thirty trials went well.

"You ignore everything known before the pilot." True. The calculation treats the thirty runs as the only evidence. A supplier with thousands of logged hours on similar equipment brings real information. Ask for it with its denominators, its reference method and its misses. If it arrives in that form, weigh it. In our reading of the published literature it rarely does. A 2023 mapping review kept 24 of 392 records on digital safety interventions in construction and found low to no evidence of better working conditions or fewer injuries.

"One miss in three hundred is still a death." Yes. That is why detection is the wrong place to put the whole weight of protection, and why the first question is whether the exposure can be removed.

What to do with this

Decide the claim first. Ask the technical authority what miss rate would be tolerable for an advisory function at this exposure. Read the number of trials from the list above.

Sign the pass rule before the first run. "No miss and no late alert in 59 independent trials across three shifts" is a rule. "Performance will be reviewed" is not.

Vary the trials. Different approaches, light, operators, shifts. Record the conditions for each.

Restart on a miss. One miss stops the challenge. Find the cause, correct it and begin the count again. Do not average it away.

Say what the pilot allows. The decision paper should carry one sentence of this form: "With 59 clean trials we can say the miss rate is probably below 5% in the conditions tested."

What would change our view

Evidence that staged trials with surrogates predict performance in real work for a given function. With that, fewer trials would carry more weight. We have not found it.

A clean pilot is good news. It is also less news than it looks.

© 2026 Riskopilot · riskopilot.com · contact@riskopilot.com · CC BY-ND 4.0

Read the full manual. *Prove it on site: a field manual for piloting safety technology against critical controls* is published by Riskopilot Research on Riskolabs.

Sources

1. Hanley, J. and Lippman-Hand, A. (1983). If nothing goes wrong, is everything all right? Interpreting zero numerators. *JAMA* 249(13), 1743 to 1745. doi:10.1001/jama.1983.03330370053031.
2. Ruff, T. (2006). Evaluation of a radar-based proximity warning system for off-highway dump trucks. *Accident Analysis & Prevention* 38(1), 92 to 98. CDC Stacks. One truck, one mine, seven days.
3. Trask, C. and Linderoth, H. (2023). Digital technologies in construction: a systematic mapping review of evidence for improved occupational health and safety. *Journal of Building Engineering* 80, 108082. ScienceDirect.

The trial numbers are a Riskopilot calculation, 1 – 0.05^(1/n), one-sided at 95%. This article proposes no acceptable failure probability. The site's technical authority sets it.

Published under CC BY-ND 4.0. Share unchanged, with attribution. General information, not engineering or legal advice.